STUDIES IN MATHEMATICAL ECONOMICS AND ECONOMETRICS

THE UNIVERSITY OF CHICAGO PRESS
CHICAGO, ILLINOIS

THE BAKER & TAYLOR COMPANY
NEW YORK

THE CAMBRIDGE UNIVERSITY PRESS
LONDON

THE MARUZEN-KABUSHIKI-KAISHA
TOKYO, OSAKA, KYOTO, FUKUOKA, SENDAI

THE COMMERCIAL PRESS, LIMITED
SHANGHAI

STUDIES IN MATHEMATICAL ECONOMICS AND ECONOMETRICS

In Memory of Henry Schultz

Edited by

OSCAR LANGE · FRANCIS McINTYRE
THEODORE O. YNTEMA

THE UNIVERSITY OF CHICAGO PRESS
CHICAGO · ILLINOIS

COMPOSED AND PRINTED BY THE DENTAN PRINTING COMPANY
COLORADO SPRINGS, COLORADO


Henry Schultz

PREFACE

THE tragic death of Henry Schultz on November 26, 1938, has deprived economic science of one of its most outstanding representatives. The Department of Economics of the University of Chicago, of which Henry Schultz was a member, is honoring his memory by publishing this volume of essays in the field of economic research to which he had devoted his life. The invitation to contribute extended to those who were associated with Henry Schultz, either as students or colleagues or through work in the same field, met with an overwhelmingly positive response. Unfortunately, because of the war the contributions from European authors were fewer than would otherwise have been the case. To all those who took part in this enterprise we want to express our sincerest thanks. In presenting the volume to the public, we hope that it will prove a worthy memorial to our colleague and friend.

OSCAR LANGE
FRANCIS MCINTYRE
THEODORE O. YNTEMA
Editorial Committee

TABLE OF CONTENTS

I
HENRY SCHULTZ

HENRY SCHULTZ—FRIEND[1]

GEORGE K. K. LINK
University of Chicago

IT IS my sad privilege on this occasion to speak of Henry Schultz as friend. Our friendship had its beginning in his desire to enjoy the out of doors. Six years ago, at the Round Table in the Quadrangle Club, he heard Benjamin Willier, Sewall Wright, and myself refer to our hikes in the Indiana and Michigan dunes. He expressed his interest, was invited to accompany us, and promptly became a regular participant. Later Arthur Dempster and Samuel Allison became members of the group so that it included an embryologist, a geneticist, a pathologist, an economist, and two physicists. On every trip one or more of us carried a book, a journal, or a separate in the knapsack. It usually fell to my lot to read aloud while we sat around a fire, or in warm weather enjoyed either the sun or the shade. We had very spirited discussions on a wide range of old and contemporary subjects. Henry and Sewall shared special interests in mathematics and statistics, particularly the problems of probability. Henry also became interested in all sorts of things biological, including his own soma and psyche. In the winter Benjie, Sewall, Henry, and I often walked from Tremont to Michigan City and back along the beach where the sand and snow offered excellent facilities for execution of biologic diagrams, curves, and equations of all sorts. Henry and I, alone or in company with one or more of the other members of the group, made trips to the dunes in all sorts of weather throughout the year for almost six years.

Henry was a good companion. He was an excellent walker, was physically tough and wiry, and above all was uniformly good natured, even in the worst of weather. Even though he lacked manual aptitude and experience, he was always eager to do his bit. Reared as a metropolitan city lad, he had had no out-of-door life before his experiences as private and sergeant in France during the World War. His army equipment, bleached, worn, and battered, served as his early dune outfit. He derived great pleasure from gradual acquisition of new pieces of clothing and of other paraphernalia as his finances per-

[1] Read with minor omissions at the Henry Schultz Memorial Service held at the University of Chicago, January 30, 1939.

his boyish delight, and his new gadgets were the subject of much mitted from time to time. The rest of us had lots of fun with him in good-natured banter. Two years ago he outdid himself and the rest in the acquisition of the latest Bergens Rücksack. Few boys, little or big, have gotten more satisfaction out of putting over something than did Henry in secretly getting a bag that excelled every other one of the group. Those to whom its virtues had been revealed were pledged to secrecy so that he might have the pleasure of surprising the others on successive trips. Each time, kneeling beside the bag, glancing up at us, his black eyes shining with pride and pleasure, Henry unveiled and operated the swivel support, the automatic closing cord, and the zipper pockets. He got as much satisfaction out of that bag, I believe, as he did from his prize apparatus, the harmonic analyzer, which to us was known as the "hormonic analyzer" and by us was held responsible for the fact that marriage deprived him of one assistant after another.

His last acquisition to be shared with us was a tartan shirt. In 1937, after the manuscript of Henry's book had gone to the printer, he and his wife realized a long-cherished ambition to go to the Rocky Mountains. After a few days spent in Glacier Park, Montana, Henry conceived the idea of surprising Mrs. Link and me at our favorite camp in the Canadian Rockies. They arrived in Banff after a series of hard bus rides, only to be informed that the camp had closed that day, but that we probably could be found at Lake Louise. There they surprised us by walking in while we were at supper. For Henry there was bitter with the sweet. In Banff he acquired a tartan shirt, the most chromatic available, in which he had hoped to outshine me, but unfortunately he found me wearing one which made his look anemic in comparison.

Fortunately, the weather was ideal during Henry's two-day stay in the Canadian Rockies, and he was full of the zest of adventure. Mrs. Link and I had the pleasure of sharing with him and Mrs. Schultz mountain scenery such as they had not seen before. The first day was devoted to going to the head of the Bow Valley. From Bow Pass, Henry had his farthermost look into the North Country. The next day we made an eighteen-mile hike and climb into Paradise Valley. We got above timber line, above the larches which looked like spun gold in the September sun. Facing the great glacier which nestles at the foot of the immense cirque of Mount Hungabee, we had a bath in the hot sun with cooling breezes from the glacier and from the waterfalls that leap thousands of feet to the talus below. Never before had I seen him so happy nor so relaxed. He spoke of this experience many

times during the past year. It was obvious that he had caught the spell of the western mountains.

Henry was an interesting and a stimulating companion. Professionally he was engaged in the application of mathematics to problems in the field of economics. He was interested in the theoretical and in the quantitative aspect of things, but this did not make him contemptuous of the empirical, the qualitative, and the historical. He practiced both analysis and synthesis. Though a staunch champion of the use of concepts and propositions in science, he was not greatly interested in metaphysics. While he was aware that there can be no civilization "apart from metaphysical presupposition," he considered much of metaphysics, especially the dialectical, an interesting intellectual sport, which, he thought, Santayana aptly described as "a futile attempt to determine matters of fact by means of logical or moral or rhetorical constructions." While Henry sought to give concepts meaning through operational procedure, he nonetheless agreed with Santayana that "the notions of the scientist are so remote from the scale and scope of our senses that they can't be anything but schematic, full of abstractness and vagueness." The economic laws which he sought to formulate and to express mathematically, he was aware, are empirical extrapolations of present situations, devoid of certainty and, at best, only probability statements.

In "Retrospect and Prospect," the last chapter of his book, he gives us a glimpse of the inner light that led and sustained him in the scientific toil to which he devoted all of his time and energy. "The most," he wrote, "that any one worker can do is, of course, to break up the main task into its component parts and attack as many of these as he can in his lifetime. He must not get discouraged before the enormity of the task, for as Aristotle tells us, 'The search for truth is in one way hard and in another way easy, for it is evident that no one can master it fully nor miss it wholly. But each adds a little to our knowledge of Nature, and from all the facts assembled there arises a certain grandeur.' "

While Henry emphasized the role of reason in adding to and assembling our knowledge of Nature, his conception of reason was naturalistic. He considered it a product of phylogenetic, ontogenetic, and social development of organisms, not a supernatural entity which establishes a temporary and hectic domicile in a body which it despises and considers an impediment to its essential activities and a contaminant of its purity. He was of the opinion that what often is called reason is in fact rationalization of emotions, intuitions, appetites, hopes, and fears. He believed that disorder is as deep seated as

order and that the irrational is a basic ingredient of the world. He was aware that man does not live by reason alone but even more by faiths whose common denominator is Santayana's "animal faith." He also was aware that man, unlike other animals, is able to metamorphose tenets of faith into articles of truth, and therein has at his disposal means, not granted to other animals, for deceiving himself and for developing arrogance and cruelty. Henry was pragmatic and pluralistic and utterly out of sympathy with the doctrine of dogmatic finality. He was opposed to authoritarianism, philosophical, theological, moral, ethical, and religious, and to political and economic totalitarianisms.

Henry practiced severe self-criticism. Toward the work of others he was a keen but not an unkindly critic. Unkindness in others he charitably attributed to mental and physical illness.

He means a great deal to me because of a series of discussions we had on the problem of norms, including human values and ideals. Far apart as the fields of economics and pathology may seem, we soon found that they share common ground in many basic concepts. When we embarked on our discussions about utility, a basic concept in economics, and about injury, injury reaction, disease, and death, basic concepts of pathology, we had no intention of landing on the general problem of evil, which is the rock upon which metaphysics and theology generally suffer shipwreck. Nevertheless, we did land there, and we explored it.

We rejected the concept of absolute values and the notion that the universe aims at one type of perfection. We were in agreement on the proposition that most evil at the level of consciousness comes from awareness of the difference between ends sought or entertained and ends realized and from man's reactions to the resultant sense of frustration, of failure, of tragedy. We were aware that this lack or loss of harmony arises in part out of conflicts of values either within the same individual or between individuals, or between groups, or between individuals and groups. We could not escape the proposition that such disharmony is inevitable in a world of diversity and change —a world in which generation, growth, development, or persistence of one thing depends upon cessation, destruction, decrease, or retrogression of another; a world in which, in short, realization of one possibility often excludes realization of others. Nor could we ignore the proposition that so long as spontaneity indicates new aims in the process of creative adventure, there will be discord, confusion, conflict.

We were aware that imperfect realization of ends often is the source of furious activities of individuals and of groups; that avoid-

ance of frustration by nonentertainment of ends solely because they may not be realized leads to decadence; and that few individuals and social groups are able to believe they are getting the better of frustration by proving they can stand worse.

We shared the hope that the conflict of values might be made as harmonious as possible through greater practice of patience, good will, tolerance, humility, and compassion. We eventually agreed that for attainment of the greatest harmony possible in dynamic creative living with its inevitable imperfect realization of ends, the sense of tragedy probably needs be accompanied by what Whitehead[2] calls "the intuition of peace" and defines as "the inner feeling which belongs to the grasp of the service of tragedy." We considered this quality of mind essential because it "keeps vivid the sensitiveness to tragedy while it sees tragedy as a living agent persuading the world to aim at fineness beyond the faded level of surrounding fact." This feeling of peace amid pain, frustration, loss, tragedy, is needed, we felt, because it permits continued aim at social perfection in that it leads to "conformation of purpose to ideals beyond personal limitations" and enables the wise man to face his fate, to master his soul.

The site of most of these discussions was a high dune capped with some open spots, a cluster of linden and cherry with wild grape and bittersweet, and patches of sand grass, leading a perilous existence in their contest with the sand and wind. There were days of dazzlingly blue sky; days with brisk winds and rapidly moving clouds that cast fleeting shadows on water, sand, and forest; days with overcast skies, mists, fog, rain, snow, and ice. From our site we could look north to the meeting of lake and sky, and to the south across moving sand and forest to the cloud-breeding moraine, a massive, mute reminder of the front of a continental ice sheet and of a lake whose surface once stood above us. Below, in the bare sand, stood barkless, ghostly trunks of huge white pine, once monarchs of a forest, then buried by the sand, and now exhumed by the wind and used by bluebirds as sites for song and nesting. In the middle distance the front of a dune was advancing into a marsh. Its relentless march was obliterating a long-established habitat of plant and animal life, as well as remnants of the old Fort Dearborn–Fort Detroit trail which skirted the marsh in the days of the Indian, of the trapper, and of the trader. We often heard the call and song of bluebird, redbird, and bobwhite, and in the declining days of summer the strident chorus of the cicada. In the early spring we enjoyed the heavy scent of the

[2] A. N. Whitehead, *Adventures of Ideas* (New York: Macmillan Co., 1933), p. 392.

sand cherry, and in May, the delicate fragrance of the wild grape. In June the odor of the flowers of basswood delighted us and attracted myriads of bees that both destroyed and increased life as they pillaged and pollinated the blossoms. During October and November, when the dune was aglow with foliage of oaks, sand cherry, bittersweet, and red osier dogwood, the air was permeated with odors of overripe fruit of the wild and escaped grape. Then came winter, odorless but colorful; stark, dark frames of bush and tree with varicolored sleeping buds set off against the pinkish highlights, the blue middle tones and the purplish shadows of sand, snow, and ice.

It was here that Henry struggled with and overcame the attack of low spirits mentioned in the Preface of his book. Here we read and discussed many books. Here, too, we spent many carefree moments in enjoyment of the natural beauty of the region, of the life-giving rays of the sun which warmed and tanned our bodies, of lapsing into sleep and of awakening.

And now this venture in companionship has come to an untimely end. "In the fullness of his strength and beauty his life went out like a falling star. His spirit fled to the aether."[3]

Memory of him and of those days of companionship even now is shrouded in fantasy. Confronted with the tragedy of his death, we who experienced the joy of knowing Henry as friend find peace in the realization that the effects of his companionship continue in us and from us extend to others.

In the light of our association it seems fitting that I read passages from Jastrow's translations of the Books of Job and of Ecclesiastes reminiscent of our discussions and appropriate to the mode of his death and to this occasion.[4]

> Man is born of woman,
> Few of days and full of trouble.
> He comes forth like a flower and withers:
> Fleeing as a disappearing shadow.
>
> Where wast thou when I laid the earth's foundation?
> Declare, if thou hast the knowledge!

I have observed every ambition which God[5] has given the children of men for their affliction. He has given them a grasp of the whole world, without, however, the possibility on the part of man to fathom the work which God has made from the beginning to the end.

[3] A fragment from Euripides.

[4] These passages are taken from Morris Jastrow, Jr., *The Book of Job*, and *A Gentle Cynic, Being the Book of Ecclesiastes*. The publisher, J. B. Lippincott Co., has kindly given permission to use them.

[5] In this address the term God is used in the sense of all that which will not have been understood when the last rational being has passed out of existence.

And once more I considered all the oppressions practiced under the sun, and Oh, the tears of the oppressed without one to console them, and the violence of their oppressors with no comforter in sight!

Folly is often placed in high positions while the choice spirits sit in a humble place.

Do not say, why is it that the former days were better than these, for it is not out of wisdom that thou puttest this question.

> I experienced under the sun that
> The race is not to the swift,
> Nor the battle to the strong;
> Wise men lack an income,
> Prophets do not possess riches,
> And the learned lack wealth.
> But time and chance overtake them all.

Furthermore, man does not know his time. As fish are caught in a net, and as birds are trapped, so the children of men are entrapped at an unlucky moment, when evil comes suddenly upon them.

Henry was especially fond of this passage, often quoted the line, "But time and chance overtake them all," and for a long time considered using it in his book.

He was not able to share the exultant cry, "All's right with the world," nor like Job's friends to "undertake to defend God by a falsehood." Reflecting upon the injustices arising from rational and ideal-pursuing existence, he took his stand with Job who refuted the pious and shallow argument that apparently undeserved suffering in fact is deserved punishment imposed by an omnipotent and a just God, and declared

> Aye though he slay me, I tremble not;
> For all that, I will maintain His course to His face.

Fortunately, Henry was not overwhelmed by the tragedy of life. He agreed with Koheleth that "Light is sweet, and it is pleasant for the eyes to see the sun." So far as possible he lived in accord with Koheleth's advice:

> Therefore go, eat thy bread with joy,
> and drink thy wine with a merry heart,
> For God has already given his approval to thy deeds.
> At all times be thy garments white,
> And let oil not be lacking for thy head.
> Enjoy life with the woman of thy love,
> All the days of thy vain life,
> Which God has given thee under the sun,

for that is thy portion in life, and the compensation for thy toil under the sun. Whatever thou canst afford with thy substance do, for there is no activity, or reckoning, or knowledge, and no wisdom in Sheol,[6] whither thou goest.

[6] The term "Sheol" is used in this address in the sense of biologic dissolution.

Henry not only enjoyed a measure of these compensations but also experienced the joy of creative functioning which arises from entertainment of ideas and comes to expression in active contribution to man's vision of that "degree of orderliness which the world exhibits."[7]

In closing, let me combine an utterance of Euripides[8] with a passage from the *Iliad*:

> Zeus, whether thou art Compulsion of Nature, or Intelligence of Mankind, to thee we pray:
>
> Make clear day, and vouchsafe unto us to see. So it be but in light, destroy us.

[7] Whitehead, *op. cit.*

[8] *Trojan Women.*

HENRY SCHULTZ: HIS CONTRIBUTIONS TO ECONOMICS AND STATISTICS [1]

THEODORE O. YNTEMA
University of Chicago and the Cowles Commission

IN JANUARY, 1937, Henry Schultz delivered a public lecture before the Division of the Social Sciences of the University of Chicago on "The Quantitative Method with Special Reference to Economic Inquiry." The concluding section of that lecture he considered significant, for he later reproduced it in full in an abstract published in the *Report* of the Cowles Commission Research Conference.

> Although a theory of dynamic economics is still a thing of the future, we must not be satisfied with the *status quo* in statistical research. Many of the statistical researches that are carried on in the social sciences lack the inspiration of any theory—static or dynamic. Relatively few of us like to perform the necessary *mental experiments* which should precede the assembling and the manipulating of the data, if the results are to be significant. Relatively few of us pause to ask ourselves, *before embarking on the investigation:* What are we trying to prove or disprove? What propositions will have to be established for proving or disproving our thesis? Can they be derived from the statistics which we propose to collect? If so, by what method of analysis, and to what degree of probability? It is not surprising, therefore, that much of what passes as research in the social sciences is a more or less aimless distribution of questionnaires and a more or less mechanical elaboration of erroneous data through the use of correlation coefficients and other statistical devices. Research is not good simply because it is mathematical or statistical, or because it makes use of ingenious machines. Research is good if it is significant, if it is fruitful, if it is consistent with established principles, or if it helps to overthrow erroneous principles.[2]

The fact that Henry Schultz oriented his own research by that philosophy explains in part the abundant fruitfulness of his efforts.

He was a scholar with breadth and depth of learning. An earnest student of the great masters, Laplace, Gauss, Pearson, Cournot, Walras, and Pareto, he made them live again in his own work. He was a conservator of knowledge, consolidating past accomplishments and paving the way for future progress. He was a systematic and pains-

[1] Read at the Henry Schultz Memorial Service held at the University of Chicago, January 30, 1939. Reprinted from the *Journal of Political Economy*, Vol. XLVII, No. 2 (April, 1939).

[2] "Statistics in Economics," *Report of Fourth Annual Research Conference on Economics and Statistics* (Colorado Springs, Colo., 1938), p. 84.

taking worker, with notes and files carefully organized to preserve his accumulations of data and theory.

His various activities were integral parts of a unified plan. Statistical economics was his major interest, and for this he mastered the methods of mathematical economics and statistics and such pure mathematics as he required. His program for research showed structure: for fifteen years he labored toward the objective of measuring demand and supply functions—building the theoretical framework, perfecting the statistical methods, and extending the scope of his reading and observation. Not once did he turn aside to the irrelevant or immaterial.

By the death of Henry Schultz we lost a friend, a gentleman, an inspiring teacher, and a distinguished scholar. His life was cut short before his intellectual powers had reached their maximum or his years of study had come to full fruition. Providentially, though, we have his book, which preserves for us a broad foundation for the comprehension of problems in his field.

In reviewing Schultz's work it is convenient to consider separately his contributions in statistics, in mathematical economics, and in statistical economics.

In the field of statistics his publications dealt primarily with curve-fitting, sampling errors, and correlation methods. These were, for the most part, by-products of his major studies in statistical economics.

In curve-fitting he was concerned with the problem of finding the most probable function when more than one variable is subject to error. He brought together widely scattered references on this topic, made the procedures available in convenient form for other students, and demonstrated their significance in the determination of demand curves.[3]

In discussions of sampling errors Schultz drew heavily on the classical literature of least squares and helped place some of the recent work in its proper historical setting. His original contribution consisted in deriving the standard error of the function for several curves commonly used as time-series trends.[4]

"Elements of Curve-fitting and Correlation," which appears as Appendix C in *The Theory and Measurement of Demand,* is an elegant statement of correlation theory and practice. Because of its value to

[3] *Statistical Laws of Demand and Supply* (Chicago, 1928), pp. 38–40, and *The Theory and Measurement of Demand* (Chicago, 1938), pp. 715–17.

[4] "The Standard Error of a Forecast from a Curve," *Journal of the American Statistical Association,* June, 1930, pp. 139–85.

advanced students and research workers, this appendix was reprinted and is available as a separate monograph.

His last article appeared posthumously in the current issue of *Econometrica*, "A Misunderstanding in Index Theory: The True Konüs Condition on Cost-of-Living Index Numbers and Its Limitations." [5]

But Schultz's greatest contribution in statistics lay in his graduate teaching and in his training of assistants. Qualified by his instruction and stimulated by contact with real research, several of his students have already published excellent papers in theoretical and applied statistics.

The separation of Schultz's work in mathematical economics from that in statistical economics is rather arbitrary since they were often so closely related, the former serving as a basis for the latter.

In the mathematical formulations of economics, Schultz was a disciple of Cournot, Walras, and Pareto. The influence of Walras and Pareto has been and will be much the greater because Schultz understood and disseminated their ideas. From his first article on the "Elasticity of Demand and the Coefficient of Correlation"[6] to his book in 1938 [7] he made constant reference to the scholars of Lausanne. Most important of Schultz's own writings in mathematical economics are chapters i and iv in *Statistical Laws of Demand and Supply with Special Application to Sugar,* his article "Marginal Productivity and the General Pricing Process,"[8] and, overshadowing the rest, chapters i, xviii, and xix in *The Theory and Measurement of Demand.*

These chapters in his recent book "give a simple, yet quite adequate, summary of the general theory of utility, exchange, and demand and its latest developments—something not to be found in existing treatises." They also "present the modern theory of demand for completing and competing goods" and "develop a measure of the extent to which human behavior in the market place is rational or consistent."[9] These are Schultz's own words. We should add that these chapters are not only unique in scope but that they are also distinguished for elegance of presentation and critical discrimination. Without question they constitute the best discussion of demand extant—one that will long be classic.

Statistics and mathematical economics were, however, merely contributory to Schultz's main objectives in statistical economics.

[5] January, 1939, pp. 1–9.
[6] *Quarterly Journal of Economics,* November, 1923, pp. 169–71.
[7] *The Theory and Measurement of Demand.*
[8] *Journal of Political Economy,* October, 1929, pp. 505–51.
[9] *The Theory and Measurement of Demand,* p. x.

Henry Ludwell Moore, to whom *The Theory and Measurement of Demand* is dedicated, was the initial source of Schultz's inspiration in this undertaking. Although Schultz's own explorations surpassed Moore's in theoretical significance and quantitative accuracy, he never lost his reverence for his teacher's pioneer accomplishments.

Economic theory is a structure composed of inequalities, in which deduction is based on the signs of the first and second derivatives of various functions descriptive of utility, demand, and supply. We do not know the mathematical forms of these functions or their parameters. To the determination of approximations to these functions Schultz directed his efforts.

For the last hundred years such inductive investigation was urged repeatedly by Cournot, Jevons, Marshall, and Pareto, but almost never was it seriously attempted. As Schultz said: "The history of the statistical study of demand is essentially a recital of neglected opportunities crowned by a few belated and partial successes."[10]

Although there had been a few sporadic attempts earlier, it was not until 1914 that the first definitive and systematic attack on the problem of deriving demand curves was completed. In that year Henry Ludwell Moore published his *Economic Cycles: Their Law and Cause* containing equations expressing the relation between quantity and price for each of several agricultural products and, derived from these equations, elasticities of demand for each crop. This work has served as a point of departure for practically all the demand studies which have subsequently been made.

Since 1925, when his first major article, "The Statistical Law of Demand as Illustrated by the Demand for Sugar," appeared in the *Journal of Political Economy*,[11] Schultz was the leader in this field, increasing his eminence with each successive publication. It is not necessary, however, to review his individual articles, since practically all of them are integrated in his two books, *Statistical Laws of Demand and Supply with Special Application to Sugar* and *The Theory and Measurement of Demand.* The first of these is devoted to the determination of the demand and supply functions and elasticities for sugar, with special attention to the logical basis for procedure and with an illustrative application to the tariff. *The Theory and Measurement of Demand* is the product of a ten-year program of statistical research. In my opinion the book is great primarily because of its theoretical and procedural contributions; but the statistical chapters are also monumental achievements, dwarfing previous studies in the

[10] *Ibid.*, p. 657.
[11] October, 1925, pp. 481–504; and December, 1925, pp. 577–637.

field. I cannot better indicate the scope of the book than by borrowing the author's words.

> The present work is the first systematic treatise on the subject. It begins with a résumé of the modern mathematical theory of demand and then proceeds to summarize, compare, and evaluate the various methods and procedures that have been suggested for deriving demand curves from statistics (Part I). Finally, it utilizes some of these methods and procedures in deriving the American demand functions for sugar, corn, cotton, hay, wheat, potatoes, oats, barley, rye, buckwheat, and for beef, pork, and mutton; and the Canadian demand functions for sugar, tea, and coffee (Parts II and III). An important feature of this work is the comparison of the elasticities of demand and the rates of shift of the demand curves of the various commodities obtained by the different methods.
>
> The keystone of this treatise is the "Law of Rational Consumer Behavior" . . . . [which] frees the statistical study of demand from much of its former empiricism and gives it greater significance and utility.[12]

Great as were his statistical labors, Schultz confined them to a limited group of commodities. Consequently, more interest attaches to his methods than to his quantitative results. To study the correlation between price and quantity series, Schultz, like Moore, employed the methods of link relatives, ratios to trend, and time regression, i.e., multiple correlation, with time as one of the independent variables. From all three methods he obtained fairly consistent results, although the link-relative findings showed slightly erratic behavior. Of the three procedures, he gave greatest attention and emphasis to that of multiple correlation.

In the process of fitting the demand curve to the data, the assumption regarding the relative magnitude of the errors in the respective variables affects the results significantly. This Schultz discovered and emphasized. In his early work on the demand for sugar he applied the method of least squares on the assumption that both quantity and price were subject to error, but in his later investigations presented in *The Theory and Measurement of Demand* he abandoned this more complicated procedure. Instead, he assumed first that price alone was subject to error and then that quantity alone was subject to error and determined the demand function separately on the basis of each of these assumptions. Preferring the former because the consumption data were less reliable than the price data, he based his inferences concerning demand functions, elasticities, and errors of sampling on equations in which quantity was the dependent variable subject to error and price an independent variable not subject to error. This procedure, however, neglects the fact that "errors" arise because of shifts in the demand function not removed by deflation or time adjust-

12 *The Theory and Measurement of Demand*, pp. 659–60.

ments, as well as because of imperfections in the measurements. A full consideration of this matter would lead to some upward revision in the most probable values of the elasticities.

Schultz made extensive experiments to compare the results based on adjusted data with those based on the unadjusted figures. His adjustments involved dividing consumption by population and price by the Bureau of Labor Statistics' Index of Wholesale Prices. In passing, he recognized the possible imperfection of such deflation, but he did not stop to investigate the matter. Had he done so, he might have found the explanation for the perplexing negative partial correlation between sugar consumption and business activity,[13] but his statistical findings probably would not have been much improved by a more elaborate deflation technique.

It is curious, however, that the problem did not greatly interest him, because it is so fundamentally related to income elasticities and to other changes in buying behavior at different stages of the cycle and especially because demand studies cannot be extended far into the field of nonagricultural commodities without refinement in the deflation process. Schultz's researches were restricted to commodities characterized by relatively stable demand and, in most cases, by relatively widely fluctuating supply. The commodities possessing these characteristics are almost exclusively farm products. If industrial commodities with less erratic fluctuations in supply are to be studied, a more precise method of adjusting or deflating to compensate for shifts in the demand function will be necessary.

The finest memorial to Henry Schultz will be the continuation of his program for research in demand and supply. That program should include not only the new fields for study mapped out in the chapter "Retrospect and Prospect" in *The Theory and Measurement of Demand* but also the methodological problems of curve-fitting and deflation.

Schultz undertook to bridge the gap between factless theory and theoryless fact. He sought to integrate economic theory, which runs only in terms of inequalities, with the statistics of market price and quantity, and from these latter to discover the functions without which we can predict only the sign, but never the magnitude, of the effect resulting from a disturbance to an economic system. In that work he was notably successful. Future analysis may lead to some revision of his quantitative findings, but Schultz has established beyond question the approximate values of important economic parameters.

[13] *Ibid.*, pp. 220–21.

It is difficult for one who has not read *The Theory and Measurement of Demand* to appreciate the magnitude of the author's achievement in the accumulation, organization, and exposition of so vast a fund of knowledge. For fifteen years Schultz labored to gather material for his magnum opus, drawing together from economic theory, mathematics, and statistics all that provided the framework for his hypotheses or the means for their verification. Thus he created a classic which will take its place with the works of Cournot, Walras, and Pareto whom he so revered.

His was the creed of the scientist. The book ends with a quotation from Aristotle:

> The search for truth is in one way hard and in another easy, for it is evident that no one can master it fully or miss it wholly. But each adds a little to our knowledge of Nature, and from all the facts assembled there arises a certain grandeur.[14]

[14] *Ibid.*, p. 666.

HENRI SCHULTZ, PIONNIER DE L'ECONOMETRIE

RENÉ ROY

Ecole Nationale des Ponts et Chaussées,
Institut de Statistique de l'Université de Paris

C'EST à Paris, vers 1934, que nous avons fait la connaissance d' Henri Schultz; nous l'avons rencontré à nouveau en 1938, au Congrès de la Commission Cowles à Colorado Springs. Ayant eu ainsi l'occasion de l'apprécier, nous avons éprouvé une sincère émotion à la nouvelle de sa fin tragique et prématurée. Nous avons le sentiment qu'avec lui disparaît un des meilleurs serviteurs de la Science Economique au moment même où en possession de toute sa maîtrise et d'une expérience acquise par vingt années de travail, il aurait été en mesure d'accroître encore ses importantes contributions aux progrès de cette science. Aussi, nous félicitons-nous de pouvoir collaborer à la diffusion de son œuvre parmi ceux qu'intéressent les applications des mathématiques et de la statistique à la matière économique.

Si nous avons choisi pour ce bref essai le titre "Henri Schultz, Pionnier de l'Econométrie," c'est que nous considérons cet auteur comme vraiment représentatif des tendances qui doivent guider les adeptes de l'Econométrie: établissant un pont entre les élaborations spéculatives de la vieille école mathématique et celles de l'école historique, soucieuse avant tout de description, l'économétrie se propose d' atteindre le mesurable en applicant systématiquement l'observation statistique aux éléments que les théories de l'école mathématique ont permis de mettre en lumière et de relier par des équations de mutuelle dépendance. Or, telle est précisément l'attitude prise par Henri Schultz dans les recherches qu'il a effectuées pendant une quinzaine d'années, car le souci de garder le contact avec la théorie tient dans ses œuvres une place aussi grande que la mesure des éléments numériques. Il insiste au reste maintes fois sur l'obligation pour les économètres de s'inspirer constamment des théories édifiées par l'économie pure. Telle est sans doute la raison pour laquelle notre attention fut vite attirée vers les travaux d'Henri Schultz; nous devons toutefois ajouter que cet attrait fut encore accru par l'analogie de nos champs d'investigation. Henri Schultz a plus particulièrement dirigé ses efforts vers les lois de la demande et la détermination des coefficients d'élasticité; ce

domaine ayant retenu notre attention, il était tout naturel de songer à confronter nos méthodes et nos résultats avec ceux de cet auteur.

A notre sens, les contributions les plus originales d' Henri Schultz concernent:

En premier lieu, l'élasticité de la demande et le déplacement temporel des courbes de demande pour un certain nombre de produits agricoles (sucre, mais, céréales, fourrages, etc.).

En second lieu, les relations d'interdépendance pour la demande des biens complémentaires ou concurrents.

I. ÉLASTICITÉ DE LA DEMANDE

Pour ce qui regarde le premier de ces objets, c'est-à-dire l'élasticité de la demande et son déplacement temporel, Henri Schultz, s'il a enrichi par d'importantes et multiples applications nouvelles les méthodes imaginées par M. Henri Moore (chaînes relatives et rapports à la tendance) a systématiquement recouru à une autre méthode qui semble bien lui être personnelle, et qui, par sa forme même, traduit une conception dynamique de la demande: nous voulons parler de la régression temporelle. Cette dernière méthode, qui exprime le prix ou le débit d'un article en fonction explicite du temps et d'autres paramètres (prix, quantités, revenus, indices des prix ou de l'activité économique, etc.) se relie en effet, non seulement aux conceptions mères de Cournot et de Marshall sur la demande et son élasticité, mais encore à la notion générale d'équilibre économique. En pratique d'ailleurs, l'application de la régression temporelle à la détermination des élasticités ne fait souvent intervenir que le temps et le prix de l'article étudié. Sans nous attarder au détail de la méthode ou à ses résultats, qui sont aujourd'hui connus de tous, nous voudrions présenter ici quelques observations dictées par nos propres réflexions.

1°) Dans nos recherches personnelles, nous nous sommes essentiellement attachés à la demande des services monopolisés (transports publics, postes, tabacs) en nous bornant à invoquer la conception simple de Cournot, c'est-à-dire la loi de demande statique fondée sur la considération d'une seule variable indépendante, le prix de la marchandise ou du service étudié (tarifs).

A la faveur de modifications de tarifs, dont l'incidence est limitée dans le temps, il est ainsi possible d'apprécier l'effet qui en résulte pour le débit, et d'aboutir, moyennant une hypothèse sur la forme probable de la courbe de demande, à la détermination du coefficient d'élasticité. C'est là une méthode que nous nous croyons autorisés à qualifier d'expérimentale, puisqu'elle implique le recours à un univers

économique rigide où ne varient que deux éléments: le prix, variable indépendante et le débit, fonction de cette variable. Le choix même du champ d'investigation permet cette assimilation, puisqu'en matière de monopole, le prix est fixé par voie d'autorité. D'autre part les possibilités de substitution ou de complémentarité sont supprimées ou du moins réduites par le défaut de concurrence et par la nature particulière des marchandises ou services monopolisés.

Henri Schultz au contraire s'est hardiment aventuré dans le domaine beaucoup plus mouvant des denrées agricoles, où il n'est plus possible de recourir aux conceptions simples de Cournot et de Marshall. En un tel domaine, prix et quantités varient simultanément, et c'est toute la complexité des sciences d'observation qui domine le sujet. Force est de recourir alors aux procédés les plus délicats et les plus raffinés de l'observation par voie statistique. Pour bien marquer la distinction entre les deux méthodes, nous pensons qu'il n'est pas inutile d'exposer brièvement les réflexions suggérées par des recherches qui peuvent être regardées comme participant à la fois de ces deux méthodes. Nous aurons l'occasion d'y constater que la méthode expérimentale et la méthode de régression temporelle ont donné des résultats comparables. C'est là une constatation intéressante, car nous n'étions pas en droit de compter à priori sur une telle concordance.

2°) Les recherches auxquelles nous venons de faire allusion concernent la demande du trafic postal en France; nous avons tout d'abord effectué quelques déterminations en utilisant des procédés rudimentaires, puis à notre instigation, M. Morice a consacré une étude statistique précise à la demande du trafic postal en France et l'a publiée dans le numéro d'octobre 1938 de la revue *Econometrica.*

Au cours de la période comprise entre la fin de la guerre 1870–71 et le début de la guerre 1914-18, le tarif postal ne subit que deux modifications: (a) abaissement de 0 fr 25 à 0 fr 15 en 1878 et (b) abaissement de 0 fr 15 à 0 fr 10 en 1906.

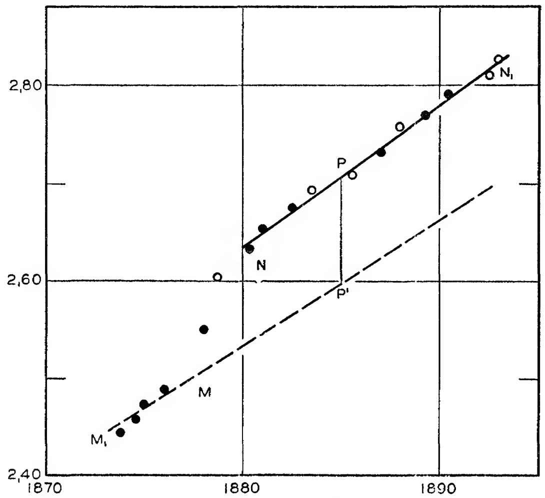
L'observation montrant que l'influence de chaque abaissement de tarifs ne porte que sur cinq ans au plus, la période étudiée constitue, en dehors de ces courts intervalles de perturbation, un champ d'investigation particulièrement propice à l'étude de la tendance séculaire; le tarif n'étant pas modifié, la nature du service en cause ne l'étant pas non plus, c'est le seul écoulement du temps qui agit sur le développement de la consommation. Or, cette tendance séculaire affecte une allure exponentielle particulièrement bien caractérisée.

Pour la période 1881-95 par exemple, où le prix n'a subi aucun changement, le nombre des lettres accuse un taux annuel moyen d'accroissement égal à 2,5%; l'écart maximum entre les valeurs observées

et les valeurs calculées est inférieur à 2%, tandis que l'écart quadratique moyen ne dépasse guère 1%.

Une telle constatation dans un domaine aussi exceptionnellement stable, justifie le choix d'une tendance séculaire à forme exponentielle, c'est-à-dire à taux d'accroissement relatif constant, pour l'application de la méthode de régression temporelle.

Quant à l'influence des deux modifications de tarifs pratiquées respectivement en 1878 et 1906, elle peut être appréciée par l'écart entre les deux droites parallèles qui, avec une échelle ordinaire pour les temps et une échelle logarithmique pour le nombre des lettres, représentent la tendance séculaire à forme exponentielle.



LOGARITHMES DU NOMBRE DE LETTRES (EN MILLIONS)

Mais s'il est possible d'apprécier la valeur absolue de cette influence, il est plus malaisé d'en déduire la valeur du coefficient d'élasticité, car nous sommes en présence de modifications importantes pour les prix comme pour les consommations, et nous ne pouvons les assimiler aux variations infiniment petites considérées pour définir le coefficient d'élasticité. Le problème ainsi posé requiert toute notre attention, car de sa solution dépend évidemment la valeur du coefficient d'élasticité répondant à une variation du tarif de tout service monopolisé ou plus généralement à toute variation discontinue de prix.

Nous résumons dans le paragraphe ci-après les réflexions que nous a suggérées le problème en question.

3°) Le coefficient d'élasticité est défini comme rapport des différentielles logarithmiques de quantités et de prix; cette définition n'est donc valable que pour de petits écarts à partir d'une certaine

position d'équilibre. Tout au plus peut-on l'étendre pratiquement à des variations finies, exprimées en pourcentages, à condition que ces variations restent limitées.

Mais la question change d'aspect lorsque nous sommes en présence de deux positions nettement distinctes, répondant à des écarts importants de prix et de quantités; chacune de ces positions ne définissant qu'un point de la courbe de demande, peut-on encore parler d'élasticité de la demande et que signifie au juste cette expression dans un tel cas?

Ce problème est à rapprocher du problème de l'élasticité d'un arc de courbe de demande, lequel a donné lieu à certains travaux (notament MM. Dalton, Allen et Lerner, *The Review of Economic Studies,* Vol. 1, n° 3 [juin 1934]). Les auteurs en question définissent l'élasticité de l'arc en se fixant à priori certaines conditions, selon un procédé qui peut être rapproché de celui qu'utilise M. Irving Fisher pour les index de prix et de quantités.

Comme l'indique M. Allen, la solution du problème ainsi posé est indéterminée, car elle dépend des conditions que l'on se fixe à priori; or, à raison même de leur empirisme, ces conditions varient nécessairement suivant les auteurs.

Ici, non seulement nous avons à définir l'élasticité de l'arc, mais nous avons à faire une hypothèse sur la forme de la courbe entre les deux points correspondant respectivement aux deux valeurs du tarif, puisque nous ne connaissons de cette courbe que les deux points en question.

Nous fondant sur les présomptions que l'on peut dégager au sujet du coefficient d'élasticité, lorsqu'on étudie les rapports entre la demande et la distribution des revenus, nous avons estimé que le plus simple et le plus conforme à la théorie, était d'admettre entre les deux points considérés l'existence d'une courbe à forme marshallienne, c'est-à-dire à élasticité constante. Nous reviendrons d'ailleurs ultérieurement sur les services que rend l'usage de cette forme particulière pour les études concernant l'élasticité de la demande.

Si nous opérons ainsi, le coefficient d'élasticité se trouve défini par la relation:

$$\lambda = \frac{\log q_2 - \log q_1}{\log p_2 - \log p_1} ,$$

p_1 et q_1 désignant les prix et quantités répondant au premier point de la courbe, p_2 et q_2 les prix et quantités répondant au second point de la courbe.

Si nous posons:

$$p_2 = p_1(1+\alpha)$$
$$q_2 = q_1(1+\beta) ,$$

nous en déduisons:

$$\lambda = \frac{\log(1+\beta)}{\log(1+\alpha)} .$$

Lorsque les variations relatives α et β sont faibles, nous pouvons nous limiter aux termes du premier ordre et prendre le rapport β/α comme valeur approchée de λ; cela revient à utiliser la définition pure et simple de l'élasticité en un point de la courbe de demande.

Si au contraire α et β sont importants, nous ne pouvons nous contenter des éléments du premier ordre; en utilisant ceux du second ordre, nous obtenons:

$$\lambda = \frac{\beta - \dfrac{\beta^2}{2}}{\alpha - \dfrac{\alpha^2}{2}} .$$

L'usage de cette formule équivaut à mesurer les variations relatives, non par rapport aux valeurs initiales des prix et des quantités, mais par rapport aux moyennes arithmétiques des valeurs initiales et finales.

Cette dernière formule est précisément celle qui a été finalement préconisée par MM. Allen et Lerner, en partant comme nous l'avons indiqué, de conditions tout à fait différentes.

Quoiqu'il en soit, nous devons retenir ici qu'il était indispensable de résoudre ce problème avant d'effectuer de plus amples déterminations du coefficient d'élasticité, toute comparaison dans le temps ou dans l'espace étant évidemment subordonnée au choix de la méthode adoptée pour le calcul des élasticités. Par sa nature même et par sa simplicité, la demande à élasticité constante s'est pour ainsi dire imposée à notre esprit comme à presque tous ceux qui ont abordé ce genre de problème.

Dans le cas particulier du trafic postal en France, cette méthode a permis la détermination des coefficients d'élasticité correspondant aux deux abaissements indiqués ci-dessus (1878 et 1906): l'élasticité, qui pouvait être évaluée à 0,60 vers 1878 s'est abaissée à 0,40 environ vers 1906; nous ne saurions trop insister sur l'intérêt que présente cette diminution du coefficient d'élasticité, caractérisant selon nous les périodes de développement économique.

4°) Lorsque M. Morice en arrive à la période 1922–36, il ne lui est plus possible de s'en tenir à la méthode expérimentale qu'il a utilisée pour la période précédente.

a) En premier lieu, les variations de tarifs se multiplient de telle manière qu'on ne peut plus disposer d'un intervalle de temps suffisant pour apprécier l'incidence de chaque modification considérée isolément.

b) En second lieu, même pour les périodes où le tarif reste constant, ce tarif ne présente plus en réalité qu'une valeur nominale, par suite des importantes variations subies par le pouvoir d'achat de la monnaie.

Ce qu'il faut alors considérer, c'est le tarif réel, quotient du tarif nominal par l'indice des prix; or, ce tarif réel ne présente aucune stabilité, à cause des fluctuations monétaires et l'on se trouve ainsi en présence de variations continues de prix, tout à fait comparables à celles qui se présentent pour les denrées courantes. Le monopole ne constitue plus dans ce cas un champ d'expérimentation exceptionnel et il est nécessaire de faire appel aux méthodes employées par l'école américaine.

L'ajustement opéré par la méthode de régression temporelle, moyennant l'emploi d'une formule à élasticité constante et à taux d'accroissement annuel également constant, conduit à des constatations dignes d'être retenues:

Tout d'abord, le coefficient d'élasticité s'établit à 0,40, c'est-à-dire à la même valeur qu'avant la guerre de 1914–18; d'autre part le taux d'accroissement annuel atteint environ 2% comme avant la guerre. L'ajustement est très satisfaisant, puisque l'écart maximum est de 3,6% et l'écart quadratique moyen de 2%.

Il importe d'insister avant tout sur l'identité des valeurs trouvées pour le coefficient d'élasticité (0,40), d'abord par la méthode expérimentale en période économique stable (1906–13) puis par la méthode de régression temporelle en période instable (1922–36). Cette identité constitue à notre sens un argument propre à justifier l'emploi de la méthode de régression temporelle qui implique le recours plus ou moins explicite à certaines hypothèses, notamment l'existence d'une surface de consommation, de forme analytique déterminée à priori, avec un nombre de paramètres extrêmement restreint.

II. DEMANDES INTERDÉPENDANTES

Relativement aux phénomènes de concurrence et de complémentarité, c'est-à-dire à l'étude générale de l'interdépendance des demandes, la position d'Henri Schultz est tout à fait significative: chacun des deux mémoires qu'il a consacrés à cette importante question débute par une analyse théorique, particulièrement consistante, puisqu'elle

s'étend jusqu'à l'origine même des conceptions qui ont engendré la notion d'équilibre économique. Ce sont donc les premières déductions de Walras et plus encore celles de Pareto, sur la fonction d'utilité, qui sont employées par Henri Schultz comme support de ses recherches statistiques. A ces notions fondamentales s'ajoutent bien entendu les apports plus récents de MM. Slutsky, Hotelling, Hicks, Allen, pour ne citer que les plus notables. Ces dernières contributions, tout en restant subordonnées à des hypothèses difficiles à contrôler, offrent à l'observateur le grand mérite de faire intervenir explicitement le revenu individuel et se prêtent également mieux que les précédentes à la mise en œuvre de l'observation statistique.[1] Henri Schultz prend au reste grand soin de préciser que son objectif n'est pas seulement de parvenir à la détermination de coefficients mesurables, mais aussi et peut-être surtout d'asseoir les théories sur des bases plus solides.

Le lien entre la théorie et l'observation est donc singulièrement accusé dans cette catégorie de recherches; la complexité du problème, le nombre élevé des paramètres en jeu, la diversité des liaisons qu'ils peuvent présenter, imposaient d'ailleurs au pionnier qui s'engageait dans cette voie, le recours systématique aux théories générales de l'équilibre économique. Toute autre méthode eût par son empirisme, même, conduit à un échec et eût été en outre fort difficile à concevoir.

La méthode et les résultats d'Henri Schultz suggèrent bien des réflexions; nous en résumons ci-après quelques-unes:

1°) La condition d'intégrabilité, exprimée sous des formes accessibles à l'observation, c'est-à-dire par les équations de Slutsky ou d'Hotelling, comporte des restrictions qui ne sont nullement négligeables:

a) Les équations de Slutsky ne valent que pour un individu; leur extension à une loi de demande collective implique certaines hypothèses qui ne peuvent être toutes vérifiées en pratique.

b) Les équations d'Hotelling, valables pour une collectivité, sup-

[1] Nous rappelons que les conditions de Slutsky s'expriment ainsi:

$$\frac{\partial x}{\partial p_y} + y\frac{\partial x}{\partial r} = \frac{\partial y}{\partial p_x} + x\frac{\partial y}{\partial r},$$

x et y désignent les quantités de 2 biens interdépendants, p_x et p_y leurs prix respectifs, r désigne le revenu de l'individu en cause.

Quant aux relations d'Hotelling elles se présentent sous la forme:

$$\frac{\partial x}{\partial p_y} = \frac{\partial y}{\partial p_x},$$

qui suppose les quantités déterminées en fonction des prix, ou sous la forme réciproque obtenue en supposant les prix déterminés en fonction des quantités.

posent que l'utilité marginale de la monnaie reste constante. Cette hypothèse, qui demeure admissible lorsque les revenus sont répartis entre un grande nombre de consommations, dont chacune n'intéresse qu'une faible partie de ses revenus, se concilient mal avec les observations faites par M. Ragnar Frisch, dans les applications de sa méthode des isoquantes.

Il n'est donc pas surprenant que la vérification statistique de ces deux types d'équations se heurte à de sérieuses difficultés, et que les résultats obtenus à ce sujet ne soient pas nettement concluants.

2°) La comparaison des divers coefficients montre que la consommation d'un article est avant tout commandée par les variations de son propre prix, l'influence des autres facteurs, et notamment celle des autres prix restant beaucoup plus faible. On observe même que la prise en considération de tous les autrex prix réduit la précision des coefficients relatifs à l'influence propre du prix de l'article étudié.

Ces constatations sont fort précieuses, car elles justifient le recours aux conceptions simples de Cournot et de Marshall, qui ne considéraient que le prix de l'article étudié. Les travaux d'Henri Schultz permettent ainsi de s'en tenir à ces conceptions, qui déterminent une approximation suffisante dans bien des cas, compte tenu de l'imprécision affectant le calcul des coefficients.

3°) Dans le même ordre d'idées, les applications font apparaître certaines contradictions[2] qu'il n'a pas été possible d'expliquer, et dans la plupart des cas, la conclusion se traduit plutôt par une présomption que par une certitude, quant à la nature des liens qui unissent les diverses demandes.

Etant donné les difficultés du problème, étant donné aussi les imperfections du matériel statistique, ces résultats ne sont pas surprenants; ils ne diminuent en aucune manière les mérites de celui qui a entrepris de telles recherches.

III. QUELQUES OBJECTIONS

En dehors des réflexions qui précèdent, certaines objections nous viennent à l'esprit.

1°) Dans les travaux concernant les applications de la statistique à la détermination des courbes de demande, il semble bien qu'on ne dispose pas toujours du matériel approprié à ce genre de recherches. Autrement dit, le matériel statistique emprunté à l'observation ne traduit pas les lois de la demande à l'état pur, car il provient aussi

[2] Le foin et l'avoine ont un type de dépendance contradictoire: un accroissement du prix de l'avoine augmente la demande de foin, tandis qu'un accroissement du prix du foin réduit la demande d'avoine.

bien des opérations de production ou d'échange que des opérations de consommation proprement dites. Ce caractère hybride est nécessairement la source d'erreurs ou de contradictions dans les résultats.

Le défaut d'homogénéité du matériel statistique n'a d'ailleurs pas échappé à Henri Schultz qui a justement considéré qu'il serait indispensable d'en améliorer la qualité.

C'est en partie pour échapper à ces inconvénients que nous avons orienté nos études personnels vers les services monopolisés, qui par leur nature même (transports publics, trafic postal, tabacs) ne prêtent pas à semblable critique. Mais en dehors des services monopolisés, il existe à cet égard d'appréciables différences d'un produit à l'autre: le sucre par exemple est beaucoup plus facile à étudier, en temps que bien de consommation directe, que d'autres denrées agricoles telles que le maïs, le coton ou les plantes fourragères.

2°) Henri Schultz porte indifféremment son choix sur des équations linéaires ou des équations à élasticité constante; il observe cependant que si les secondes donnent immédiatement la valeur de l'élasticité, les premières ne déterminent qu'une élasticité variant d'une année à l'autre, selon la place occupée dans l'échelle des prix. Les variations du coefficient d'élasticité ne sont aucunement négligeables, puisque de 1890 à 1914, ce coefficient a sensiblement varié pour la demande du sucre aux Etats-Unis entre 0,70 et 0,30.

Nous marquons au contraire notre préférence pour la demande à élasticité constante, puisque l'objet essentiel da la recherche est précisément la détermination du coefficient d'élasticité. Et puis, comment effectuer les comparaisons d'une période, d'un pays ou d'un produit à l'autre, si l'on utilise une formule à élasticité variable? Dans sa comparaison des élasticités relatives à divers produits agricoles, Henri Schultz a d'ailleurs utilisé presque uniquement des formules à élasticité constante, précisément parce que les comparaisons à effectuer imposaient ce genre de formules.

A cet avantage pratique aussi bien que théorique, s'en ajoutent bien d'autres en faveur de la formule à élasticité constante:

a) Nous pensons que si l'élasticité dépend du prix, elle peut être néanmoins regardée comme approximativement constante à une époque déterminée, pour un article donné, car la valeur de ce coefficient est en rapport étroit avec le degré d'urgence que les acheteurs attribuent à la consommation de l'article en cause.

En revanche, l'élasticité varie au cours du temps et nous pouvons même considérer la diminution de ce coefficient comme l'indice d'un régime économique progressif.

b) Ce n'est pas par hazard que le coefficient d'élasticité défini

par Marshall fait intervenir des différentielles logarithmiques et nous estimons que l'introduction de ces échelles, si fréquente aujourd'hui dans les études concernant la demande, se trouve pour ainsi dire imposée par la nature des choses.

La notion "d'effet proportionnel," clairement illustrée par certains travaux, comme ceux de M. Gibrat sur les inégalités économiques, est certainement à la base des recherches concernant les propriétés de la demande. Opérant dès lors sur des diagrammes logarithmiques, n'est-il pas logique d'admettre une forme rectiligne pour la portion de la courbe de demande intéressant les observations? A ce procédé simple et satisfaisant pour l'esprit, nous ne voyons pas d'argument qui puisse être sérieusement opposé.

3°) Nous devons insister enfin sur les contradictions observées dans les rapports entre l'élasticité de la demande et la flexibilité du prix.

L'élasticité de la demande λ et la flexibilité du prix ψ étant respectivement définies par les relations:

$$\lambda = \frac{dq}{dp}\frac{p}{q},$$

$$\psi = \frac{dp}{dq}\frac{q}{p},$$

si le prix p et la quantité q sont liés par une relation fonctionnelle, λ et ψ sont eux-mêmes tels que:

$$\lambda = \frac{1}{\psi}.$$

Mais lorsque nous opérons par voie statistique, la liaison établie entre prix et quantités est de nature stochastique et non fonctionnelle; dans ce cas, les coefficients λ et ψ résultant de l'ajustement, ne sont plus rigoureusement inverses l'un de l'autre et l'on peut constater que $1/\psi$ représente alors une limite supérieure pour λ.

La question qui se pose dans cette conjecture peut être ainsi formulée:

Doit-on réaliser l'ajustement par une expression de la forme

$$q = f(p)$$

de manière à déterminer directement l'élasticité λ ou doit-on le réaliser au moyen d'une formule du type

$$p = g(q)$$

de manière à déterminer directement ψ?

Autrement dit, est-ce l'élasticité de la demande ou est-ce au contraire la flexibilité du prix qui présente la valeur la plus significative?

En pratique, la réponse à la question ainsi posée n'est nullement indifférente: dans son magistral ouvrage *Théorie et mesure de la demande,* Henri Schultz donne pour la plupart des denrées qu'il a étudiées, les valeurs obtenues pour λ et $1/\psi$, avec le même matériel statistique; or, les écarts sont parfois considérables.

Pour le sucre par exemple, les éléments relatifs à la période 1875–1905 conduisent, avec des formules à taux d'accroissement annuel constant et à élasticité ou flexibilité constante, aux valeurs ci-après:

$$\lambda = 0{,}38\,(\pm 0{,}12)$$

$$1/\psi = 1{,}09\,(\pm 0{,}40)$$

et pourtant la demande de sucre est une de celles qui donnent les résultats les plus précis.

Pour le maïs, l'écart est moindre (environ 0,70 pour λ et 0,90 pour $1/\psi$) mais l'accord ne paraît s'établir d'une manière satisfaisante que pour la demande de pommes de terre.

Nous voyons ainsi que dans certains cas, tel celui du sucre, c'est le caractère même de la loi de la demande qui est en cause, puisque nous ne pouvons plus affirmer avec certitude que la demande est élastique ou inélastique.

Tout en mettant ces désaccords en évidence, Henri Schultz ne semble pas y attacher une très grande importance; il se contente d'observer que $1/\psi$ représente un maximum pour λ, ce qui laisse supposer que dans son esprit, l'objectif à rechercher est la détermination de λ.

D'autre part, il note à plusieurs reprises que les prix sont déterminés avec plus de précision que les quantités et il y voit un motif pour effectuer l'ajustement de la quantité en fonction du prix. De fait, l'erreur affectant le coefficient d'élasticité est presque toujours plus faible que l'erreur affectant l'inverse de la flexibilité.

Mais cet argument n'est-il pas celui d'un statisticien plutôt que celui d'un économiste? Ne peut-on soutenir en effet qu'en dehors du procédé statistique utilisé pour ajuster les données à une surface de consommation, la forme générale de l'ajustement (quantité en fonction du prix ou prix en fonction de la quantité) soit liée à la nature du problème posé dans chaque cas particulier?

Pour notre part, nous inclinons à croire que prix et quantités ne

jouent pas en général un rôle symétrique et qu'à cet égard trois cas peuvent être distingués:

a) S'il s'agit d'un service monopolisé, le prix est la variable indépendante et ce sont les variations de prix qui déterminent les variations de quantités; l'ajustement doit donc être opéré sous la forme

$$q = f(p) \ .$$

L'élément caractéristique de la loi de la demande est le coefficient d'élasticité λ.

b) Pour un marché agricole fermé, où la quantité représente la variable indépendante et le prix une fonction de cette variable, l'ajustement doit être au contraire effectué sous la forme

$$p = g(q) \ ,$$

et le facteur caractéristique est la flexibilité du prix.

c) Dans le cas d'un marché mondial, il n'est pas possible de dire à priori quel facteur joue le rôle de variable indépendante; seule une analyse de chaque marché particulier permettra de ranger le produit dans l'une des deux catégories ci-dessus.

Si les deux variables, prix et quantités, paraissent jouer un rôle symétrique, il semble que l'ajustement doive être réalisé par une relation de mutuelle dépendance répondant au type général:

$$\psi(p, q) = 0 \ .$$

Mais la difficulté n'est résolue qu'en apparence, puisqu'il est nécessaire de choisir une direction pour effectuer l'ajustement par la méthode des moindres carrés; la normale à la surface d'ajustement pourrait alors constituer une direction privilégiée.

Nous ne faisons que mentionner ces distinctions à titre de suggestion, mais nous estimons que la question ainsi posée devrait être sérieusement étudiée, sous peine de laisser planer un doute sur la valeur des méthodes et des résultats concernant les applications de la statistique aux phénomènes de consommation.

IV. CONCLUSIONS

Les objections développées dans le présent essai ne sauraient en rien diminuer notre estime pour l'œuvre d'Henri Schultz. Par son ampleur et sa forme systématique, l'investigation statistique à laquelle cet auteur a procédé constitue un élément capital de progrès et d'information que les futurs chercheurs ne pourront ignorer. Certes, des perfectionnements seront apportés aux méthodes et aux résultats,

mais dés maintenant, les acquisitions d'Henri Schultz constituent l'un des chapitres les plus riches de l'Econométrie.

On peut même raisonnablement soutenir que ces recherches ont ouvert un chapitre nouveau de la science économique, en ce sens que la méthode et le matériel statistiques ont acquis une importance que personne ne devait imaginer auparavant. C'est presque une arme nouvelle qui a été ainsi forgée et son maniement réservera sans doute bien des surprises dans un avenir qui n'est peut-être pas très éloigné.

Les adeptes de cette science se doivent donc de ne ménager à la mémoire d'Henri Schultz ni leur gratidude, ni leur sympathie.

LIST OF PUBLICATIONS OF HENRY SCHULTZ

1922. *Statistical Work of the Federal Government.* House Document 394, 67th Congress, 2d session. Also published by the U.S. Bureau of Efficiency.

1923. "Elasticity of Demand and the Coefficient of Correlation," *Quarterly Journal of Economics*, XXXVIII (November, 1923), 169–71.

1924. "The Statistical Measurement of the Elasticity of Demand for Beef," *Journal of Farm Economics*, VI, No. 3 (July, 1924), 254–78.

1925. "An Extension of the Method of Moments," *Journal of the American Statistical Association*, XX, No. 149 (June, 1925), 242–44.

"The Statistical Law of Demand as Illustrated by the Demand for Sugar," *Journal of Political Economy*, XXXIII, No. 5 (October, 1925), 481–504, and No. 6 (December, 1925), 577–637.

1927. "Cost of Production, Supply and Demand, and the Tariff," *Journal of Farm Economics*, IX, No. 2 (April, 1927), 192–209.

Review of H. A. Wallace and G. W. Snedecor, *Correlation and Machine Calculation*, in *Journal of Political Economy*, XXXV, No. 2 (April, 1927), 309–11.

"Theoretical Considerations Relating to Supply," *Journal of Political Economy*, XXXV, No. 4, (August, 1927), 437–64.

"Mathematical Economics and the Quantitative Method," *Journal of Political Economy*, XXXV, No. 5 (October, 1927), 702–6.

1928. Review of Dorothy S. Thomas, *Social Aspects of the Business Cycle*, in *Political Science Quarterly*, XLIII, No. 2 (June, 1928), 294–96.

Review of Robert Wilbur Burgess, *Introduction to the Mathematics of Statistics*, in *Journal of the American Statistical Association*, XXIII (June, 1928), 215–16.

Statistical Laws of Demand and Supply with Special Application to Sugar. Chicago: University of Chicago Press, 1928. Pp. xix + 228.

Review of Carl Snyder, *Business Cycles and Business Measurements*, in *Journal of Political Economy*, XXXVI, No. 5 (October, 1928), 637–39.

Review of G. H. Bousquet, *Vilfredo Pareto: sa vie et son œuvre*, in *Journal of Political Economy*, XXXVI, No. 6 (December, 1928), 740–42.

"Rational Economics," *American Economic Review*, XVIII, No. 4 (December, 1928), 643–48.

1929. Discussion of Holbrook Working and Harold Hotelling, "Applications of the Theory of Error to the Interpretation of Trends," in *Journal of the American Statistical Association*, XXIV, No. 165A (March, 1929), 86–90.

"Marginal Productivity and the General Pricing Process," *Journal of Political Economy*, XXXVII, No. 5 (October, 1929), 505–51.

1930. "The Standard Error of a Forecast from a Curve," *Journal of the American Statistical Association*, XXV, No. 170 (June, 1930), 139–85.

Der Sinn der statistischen Nachfragekurven, ed. Eugen Altschul. ("Veröffentlichungen der Frankfurter Gesellschaft für Konjunkturforschung," Heft 10.) Bonn, 1930. [An English translation of this, entitled *The Meaning of Statistical Demand Curves*, was planographed and distributed privately.]

1931. "Henry L. Moore's Contribution to the Statistical Law of Demand," in Stuart A. Rice (ed.), *Methods in Social Science*, pp. 644–61. Chicago: University of Chicago Press, 1931.

"The Italian School of Mathematical Economics," *Journal of Political Economy*, XXXIX, No. 1 (February, 1931), 76–85.

Discussion of G. C. Evans, "A Simple Theory of Economic Crises," in *Journal of the American Statistical Association*, XXVI, No. 173A (March, 1931), 68–72.

Review of G. C. Evans, *Mathematical Introduction to Economics*, in *Journal of the American Statistical Association*, XXVI, No. 176 (December, 1931), 484–91.

1932. "Hohe Korrelationskoeffizienten und ihre Bedeutung für das Studium der Nachfragekurven," *Allgemeines statistiches Archiv* (München), XXII, Heft 2 (1932), 293–98.

"The Shifting Demand for Selected Agricultural Commodities, 1875–1929," *Journal of Farm Economics*, XIV, No. 2 (April, 1932), 201–27.

Review of Otto Weinberger, *Mathematische Volkswirtschaftslehre*, in *Journal of Political Economy*, XL, No. 3 (June, 1932), 422–24.

"Marginal Productivity and the Lausanne School: A Reply [to J. R. Hicks]," *Economica*, XII, No. 37 (August, 1932), 285–96.

1933. Review of Jakob Marschak, *Elastizität der Nachfrage*, in *Weltwirtschaftliches Archiv*, XXXVII, Heft 1 (1933), 28–38.

"Frisch on the Measurement of Utility," *Journal of Political Economy*, XLI, No. 1 (February, 1933), 95–116.

"The Standard Error of the Coefficient of Elasticity of Demand," *Journal of the American Statistical Association*, XXVIII, No. 181 (March, 1933), 64-69.

"A Comparison of Elasticities of Demand Obtained by Different Methods," *Econometrica*, I, No. 3 (July, 1933), 274–308.

"Interrelations of Demand," *Journal of Political Economy*, XLI, No. 4 August, 1933), 468–512.

1934. "Über eine Methode zur Berechnung der Elastizität der Nachfrage und ihre Kritik durch Amoroso," *Zeitschrift für Nationalökonomie*, V, Heft 2 (1934), 216–26.

1935. "Interrelations of Demand, Price, and Income," *Journal of Political Economy*, XLIII, No. 4 (August, 1935), 433–81.

"Correct and Incorrect Methods of Determining the Effectiveness of the Tariff," *Journal of Farm Economics*, XVII, No. 4 (November, 1935), 625–41.

1936. Review of R. G. D. Allen and A. L. Bowley, *Family Expenditure: A Study of Its Variation*, in *Journal of the American Statistical Association*, XXXI, No. 195 (September, 1936), 613–17.

Review of Oskar N. Anderson, *Einführung in die mathematische Statistik*, in *Journal of Political Economy*, XLIV, No. 5 (October, 1936), 716–18.

Review of A. C. Pigou, *Protective and Preferential Import Duties*, in *Journal of Political Economy*, XLIV, No. 5 (October, 1936), 718.

1937. Review of C. J. Ratzlaff, *The Theory of Free Competition*, in *Journal of Political Economy*, XLV, No. 2 (April, 1937), 262–65.

"The Quantitative Method with Special Reference to Economic Inquiry." (A public lecture given before the Division of Social Sciences, University of Chicago, January 14, 1937.) Mimeographed.

Review of Léon Walras, *Etudes d'économie sociale* (*théorie de la répartition de la richesse sociale*) and *Etudes d'économie politique appliquée* (*théorie de la production de la richesse sociale*), in *Journal of Political Economy*, XLV, No. 5 (October, 1937), 701–2.

1938. Review of Donald H. Davenport and Frances V. Scott, *An Index to Business Indices* in *Journal of Political Economy*, XLVI, No. 2 (April, 1938), 302.

The Theory and Measurement of Demand. Chicago: University of Chicago Press, 1938. Pp. xxxi + 817.

"Mathematics in Economics," *Report of Fourth Annual Research Conference on Economics and Statistics* (Colorado Springs, Colo.: Cowles Commission for Research in Economics, 1938), pp. 77–79.

"Statistics in Economics," *Report of Fourth Annual Research Conference on Economics and Statistics* (Colorado Springs, Colo.: Cowles Commission for Research in Economics, 1938), pp. 82–84.

"Economic Theory of the Wisconsin Tariff Research Committee: A Reply to Professor W. A. Morton," *Journal of Farm Economics*, XX, No. 4 (November, 1938), 823–30.

1939. "A Misunderstanding in Index-Number Theory: The True Konüs Condition on Cost-of-Living Index Numbers and Its Limitations," *Econometrica*, VII, No. 1 (January, 1939), 1–9.

II
ECONOMIC THEORY

LÉON WALRAS' THEORY OF CAPITAL ACCUMULATION [1]

WILLIAM JAFFÉ

Northwestern University

THE object of this paper is to expound a much-neglected chapter in the history of economic theory. Special interest attaches itself at present to Léon Walras' discussions of capital, interest, and money because of the affinities which Marget[2] and Lange[3] have shown to exist between these older Walrasian theories and their independently discovered Keynesian counterparts.

What Walras actually had to say about capital accumulation is very difficult to grasp. Part of the difficulty arises from the way in which this particular discussion was incorporated into the whole structure of general static equilibrium. But even more troublesome are the obscurities and lacunae of the discussion itself.

Walras arrives at his theory of capital accumulation as follows. First he develops a pure theory of exchange, which bears a remarkable resemblance in essence to Marshall's "Temporary Equilibrium of Demand and Supply." In this part of Walras' treatise the stocks of exchangeable commodities are held constant. Then follows the pure theory of production, which corresponds approximately to Marshall's "Equilibrium of Normal Demand and Supply" in the short period. Here the commodities of the former part are viewed as products, and their quantities are subject to variations; but the quantities of productive services and, consequently, of the capital goods yielding productive services are assumed constant. It is not until we arrive at the part entitled "Théorie de la capitalisation et du credit" that the assumption of fixed quantities of capital goods and services is dropped. The whole development up to and including this point presupposes a barter or moneyless economy, for Walras' standard commodity (*numéraire*) is on a par with all other commodities or products, except that its value is taken as a standard of measurement

[1] Léon Walras, *Eléments d'économic politique pure* (4th ed. [definitive ed.]; Lausanne and Paris, 1900 [1926]), Part V.

[2] A. W. Marget, "Léon Walras and the 'Cash Balance Approach,'" *Journal of Political Economy*, XXXIX (1931), 569–600, and "The Monetary Aspects of the Walrasian System," *ibid.*, XLIII (1935), 145–86.

[3] O. Lange, "The Rate of Interest and the Optimum Propensity to Consume," *Economica* (N.S.), No. 17, pp. 20–23.

of values. Money and, therefore, interest, properly speaking, do not enter into Walras' general system until after he has evolved his theories of exchange, production, and capital accumulation without money. Monetary phenomena are then seen to affect the total equilibrium through the influence of cash balances, the demand for and offer of which must be equal if the rest of the system is to remain in a balanced state. Finally Walras discusses the conditions and consequences of economic progress. It is here that we first encounter the marginal productivity theory. In the theory of production Walras had provisionally assumed the technical coefficients fixed. Without technical progress, with fixed quantities of service-yielding factors, with a fixed population and fixed tastes, the technical coefficients might well be assumed constant in his theory of production. It would have been logically incongruous to assume anything else at that stage of the argument, for, until the theory of capital accumulation is reached, the technical coefficients in the theory of production cannot be considered as variable since this variation involves variations in the quantities of the capital goods which give rise to them. Nor could Walras have included the marginal productivity theory in his chapters on capital accumulation, for there he described only the mechanism of capital accumulation, not the reasons for it. These reasons are found in the section on economic progress. If the state of the industrial arts is unchanged, there can be no inducement to modify the technical proportions in which the productive services are combined unless population increases while the quantity of land remains fixed. When land becomes increasingly scarce, then there are inducements to change the technical proportions with a view to minimum costs, and the reasons for capital accumulation become evident.

So much for the setting of the particular discussion of capital accumulation within the total Walrasian system. For the purpose of this paper only one phase of Walras' exposition need be presented—namely, that in which Walras defines equilibrium and shows how it is mathematically *determined.* Another phase, which I am passing over in silence, is that of the *emergence* of equilibrium *ab ovo,* the egg being a random or inherited state most probably not in equilibrium.

The condition of equilibrium in production is the starting-point for Walras' consideration of capital accumulation. This equilibrium is made up of the following elements:

Given: (1) A number of individuals in a community each characterized by a group of utility functions, which are conceived in such a way that the marginal degree of utility (rareté) of each commodity is a function of the quan-

tity of that commodity only. These marginal degree of utility functions are the ϕ's in system (1) of the equations below.[4]

(2) Each individual of our community is further characterized by the resources he possesses at the moment under consideration. If there are n different types of resources (T), (P), (K), $[n=3]$, each individual has, to start with, q_t of (T), q_p of (P), and q_k of (K). These resources may be envisaged as disposal over so much flow of the corresponding services or they may equally well be envisaged as possession of so much capital goods whence these services flow. This assumption implies that the total stock of capital goods is given and constant.

(3) m different kinds of consumers' goods (A), (B), (C), $[m=3]$, (A) being the standard commodity.

(4) n different kinds of producers' goods (T) land, (P) persons; (K), capital goods in the narrow sense. $[n=3]$.

(5) nm technical coefficients $a_t\ a_p\ a_k$, $b_t\ b_p\ b_k$, $c_t\ c_p\ c_k$.

Unknown: (1) $m-1$ prices of consumers' goods: p_b, p_c, $(p_a=1)$.

(2) n prices of productive services: p_t, p_p, p_k.

(3) m quantities of consumers' goods demanded, D_a, D_b, and D_c being sums of the individual d_a's, d_b's, and d_c's.

(4) n quantities of productive services supplied, O_t, O_p, O_k being sums of the individual o_t's, o_p's, o_k's.

Such are the symbols which Walras uses. While he refrains from defining numerically the large, but finite, quantity of categories of commodities and productive services, it has been found convenient here to reduce the number in each case to manageable proportions. Now it is possible to present in brief and simple form, without sacrifice of principle, the following sets of Walrasian equations portraying equilibrium in production both for the individual and for society as a whole.

INDIVIDUAL CASE

$$\frac{\phi_t(q_t-o_t)}{p_t}=\frac{\phi_p(q_p-o_p)}{p_p}=\frac{\phi_k(q_k-o_k)}{p_k}=\frac{\phi_a(d_a)}{1}=\frac{\phi_b(d_b)}{p_b}$$

[4] The boldface symbols should be ignored until they are specifically referred to.

$$= \frac{\phi_c(d_c)}{p_c} = \frac{\phi_e(\mathbf{q}_e + \mathbf{d}_e)}{\mathbf{p}_e} \,. \tag{1}$$

$$o_t p_t + o_p p_p + o_k p_k = d_a + d_b p_b + d_c p_c + \mathbf{d}_e \mathbf{p}_e \tag{2}$$

$$o_t = f_t(p_t, p_p, p_k, p_b, p_c, p_d, \mathbf{p}_e); \quad o_p = f_p(\quad); \quad o_k = f_k(\quad) \tag{3}$$

$$d_b = f_b(\quad); \quad d_c = f_c(\quad); \quad \mathbf{d}_e = \mathbf{f}_e(\quad) \tag{4}$$

$$d_a = o_t p_t + o_p p_p + o_k p_k - (d_b p_b + d_c p_c + \mathbf{d}_e \mathbf{p}_e) \tag{5}$$

SOCIAL CASE

$$\left.\begin{array}{l} O_t = F_t(\quad); \quad O_p = F_p(\quad); \quad O_k = F_k(\quad); \\ \mathbf{D}_e \mathbf{p}_e = \mathbf{F}_e(\quad)\mathbf{p}_e = \mathbf{F}_e(\mathbf{p}_t, \mathbf{p}_p, \mathbf{p}_k, \mathbf{p}_b, \mathbf{p}_c, \mathbf{i}) \end{array}\right\} \begin{array}{c} n \text{ equations} \\ + \\ \mathbf{1\ equation} \end{array} \quad (\mathrm{I})$$

$$\left.\begin{array}{l} D_b = F_b(\quad); \quad D_c = F_c(\quad); \\ D_a = O_t p_t + O_p p_p + O_k p_k - (D_b p_b + D_c p_c + \mathbf{D}_e \mathbf{p}_e) \end{array}\right\} \begin{array}{c} m - 1 \text{ equations} \\ + \\ 1 \text{ equation} \end{array} \quad (\mathrm{II})$$

$$\left.\begin{array}{c} a_t D_a + b_t D_b + c_t D_c + \mathbf{k}_t \mathbf{D}_k = O_t \\ a_p D_a + b_p D_b + c_p D_c + \mathbf{k}_p \mathbf{D}_k = O_p \\ a_k D_a + b_k D_b + c_k D_c + \mathbf{k}_k \mathbf{D}_k = O_k \\ \mathbf{D}_k \mathbf{P}_t = \mathbf{D}_e \mathbf{p}_e = \mathbf{E} \end{array}\right\} \begin{array}{c} n \text{ equations} \\ + \\ \mathbf{1\ equation} \end{array} \quad (\mathrm{III})$$

$$\left.\begin{array}{c} a_t p_t + a_p p_p + a_k p_k = 1 \\ b_t p_t + b_p p_p + b_k p_k = p_b \\ c_t p_t + c_p p_p + c_k p_k = p_c \\ \mathbf{k}_t \mathbf{p}_t + \mathbf{k}_p \mathbf{p}_p + \mathbf{k}_k \mathbf{p}_k = \mathbf{P}_k \end{array}\right\} \begin{array}{c} m \text{ equations} \\ + \\ \mathbf{1\ equation} \end{array} \quad (\mathrm{IV})$$

$$\left.\mathbf{P}_k = \frac{\mathbf{p}_k}{\mathbf{i} + \mu + \nu}\right\} \mathbf{1\ equation} \quad (\mathrm{V})$$

Of the $2m + 2n$ equations of systems (I) to (IV) (exclusive of the boldface symbols), one can be shown to be redundant; so that we have at least one of the conditions necessary for a solution, namely, that the number of independent equations is equal to the number of unknowns.

Walras' whole theory of capitalization and capital accumulation consists in an extension, pure and simple, of this composite system in

order to include the prices and quantities exchanged of capital goods. It is my contention that, while it is perfectly possible to extend this static system to include the pricing of old capital goods, it is not possible to bring within the static structure a theory of capital accumulation without damaging both the structure and the theory.

Walras quite correctly feels that, before discussing the logic of price determination in the case of capital goods, it is necessary to establish a *raison d'être* for a capital goods market. He finds a *raison d'être* consistent with the static scheme by simply assuming that at all times some persons are eating into their capital by spending more than their incomes, while others are spending less than their incomes and buying up the spendthrifts' capital goods. Assuming this state of affairs, we may ask: What would anyone be willing to pay for a capital good (now to be distinguished from its service),[5] in view of the fact that the capital goods under consideration have no direct utility of their own, which precludes the possibility of their prices being determined on the basis of systems (I) and (II) of the production equilibrium equations (exclusive of boldface symbols)? Capital goods, if demanded at all, are wanted not for their utility, which does not exist, but for the incomes they yield; and then again not for their whole income but for their net income, these incomes being in the form of productive services. If we designate the gross income of a unit of capital good, priced in terms of the standard commodity, by p_k; the price of the capital good by P_k; the rate of depreciation as a technologically given percentage ν_k of P_k; and the rate of insurance as an actuarially given percentage μ_k of P_k, then the net income is

$$\pi_k = p_k - (\mu_k + \nu_k) P_k ,$$

from which we can obtain a rate of net income

$$i = \frac{\pi_k}{P_k} = \frac{p_k - (\mu_k + \nu_k) P_k}{P_k} .$$

By transposing, we may express the price of the capital good as

$$P_k = \frac{p_k}{i + \mu_k + \nu_k} = \frac{\pi_k}{i} ,$$

for any given i. In the case of land μ and ν are zero.

[5] It should be carefully noted that the theory of production gives the prices of the *services* of capital goods but not of the capital goods themselves. The equations of production (exclusive of the boldface symbols) on p. 40 show how the prices of even those productive services which have no direct utility are determined. Walras' "Théorie de la capitalisation" was designed particularly to enlarge the machinery of price determination in order to include the prices of capital goods as well as the prices of their services.

This is capitalization in the familiar sense of the term. While it furnishes a concept necessary for the discussion, it explains nothing, for it still does not tell us why there should be a demand for, and supply of, capital goods. If all persons in our community just lived within their gross incomes, so that there were no spendthrifts and no savers in any sense, the result would be twofold: (1) there would be no buying or selling of existing capital goods and (2) there would be no replacement of worn-out and accidentally destroyed capital goods. The net consequence would be a gradual diminution in the stock of capital goods and hence a gradual shrinkage of incomes. We should, then, have not a static state but a retrogressive state. The demand for capital goods is, therefore, a necessary condition for the existence of a static society, for dissavers are not only those who sell their capital assets in exchange for the unconsumed consumers' goods of savers but also those who fail to provide for replacements. If individuals provide for replacements, it is not because they are interested in utilityless capital goods but because they are concerned about their incomes. To keep income intact in the future the individual would have to distribute his present income, that is, $q_t p_t + q_p p_p + q_k p_k$, among: ($\alpha$) the productive services retained for his own use; (β) consumers' goods purchased; and (γ) capital replacements, as follows:

$$q_t p_t + q_p p_p + q_k p_k = \overbrace{(q_t - o_t) p_t + (q_p - o_p) p_p + (q_k - o_k) p_k}^{\alpha}$$
$$+ \overbrace{d_a + d_b p_b + d_c p_c}^{\beta} + \overbrace{q_k (\mu_k + \nu_k) P_k}^{\gamma},$$

which, when simplified, becomes

$$\underbrace{o_t p_t + o_p p_p + o_k p_k - (d_a + d_b p_b + d_c p_c)}_{e} = q_k (\mu_k + \nu_k) P_k .$$

It is curious to note that Walras appears to take it arbitrarily for granted under the blanket assumption of a progressive state that, for the community as a whole, the left member of the above equation, which he designates as e (= *excédent*), will not merely equal the right member, but will be more than equal, so that, if there are Θ individuals in our community, we have

$$\sum_{j=1}^{j=\Theta} e_j > \sum_{j=1}^{j=\Theta} q_{k,j} (\nu_k + \mu_k) P_k .$$

Neither the provision for maintenance capital nor the provision for

additional capital is as arbitrary as it seems, for it is easy enough to subsume both under the next argument which deals with the demand for more net income. In other words, if we can find a way to set up a demand function for additional net income, similar to the demand functions for anything else, the point at which the aggregate quantity demanded of new net income is zero defines the stationary state; the points at which this quantity is negative define the retrogressive state; and the points at which it is positive define the progressive state.

In the first three editions of the *Eléments* the argument did not take this course. There was no reference at all to the demand for net income. Walras simply set down an arbitrary, or what he called an empirical, gross savings function, in which the independent variables were the prices of all the products and services and also the rate of net income. He added that, if we wished to derive such a savings function from his general system, we should have to look upon utility in a new light and distinguish between present and future utilities. But actually he himself did nothing beyond asserting the possibility of such a derivation.

In his fourth edition, however, he boldly included his savings function in his total system which rests on a foundation of utility functions; but he deals here not with gross savings but with net savings which he puts in the form of what is spent on acquiring additional net income. Although he drops all reference to the difference between present and future utility, we must be careful not to interpret this as meaning that Walras repudiated or denied the element of time in his theory of savings. Given the whole structure of his general equilibrium theory, which is, as Wicksell so aptly points out, a cross-section in time of the processes involving exchange, production, capital accumulation, and circulating media, it would have been irrelevant for Walras to include any function of a lapse of time explicitly. Moreover, in a letter to Böhm-Bawerk, dated May 5, 1889, and in the Preface to the second edition of the *Eléments,* dated the same year, he indicated that he could not take the difference in value between present and future goods as a datum, for that very difference in Walras' system is a variable, functionally related to all other variables, including income. That did not preclude him, of course, from taking the preference function itself as a datum.

The precise manner in which Walras introduces his savings function into his system is this. He invents an imaginary commodity, E, which consists of perpetual annuity shares, each entitling the holder to the equivalent of one unit of standard commodity per annum. Referring back to the capitalization equation

$$i = \frac{p_k - (\mu_k + \nu_k) P_k}{p_k} ,$$

and replacing the numerator by unity and the denominator by p_e, representing the price of the income-yielding perpetual annuity, we obtain

$$i = \frac{1}{p_e} , \qquad \text{or} \qquad p_e = \frac{1}{i} .$$

This is the same thing as the formula for the discounted present value of an infinite series of net incomes. Furthermore, at the moment at which we are trying to obtain a cross-section portrayal of relationships, any individual may be regarded as possessing a certain stock of these shares, which will be numerically equivalent to his net income also expressed in terms of the standard commodity, according to the following equation:[6]

$$q_e = q_t p_t + q_p p_p + q_k \pi_k .$$

Each individual has not only a given net income at any moment of time but also a desire for net income. Although the net income is to be received in the future, the desire is felt now; and this desire will express itself now, according to the Walrasian scheme of things, as a function of the quantity of net income possessed, or $\phi_e(q_e)$. We may look at this as the marginal degree of utility function for perpetual annuity shares. And our individual, having regard to his income and the general price constellation, will decide to demand more or less of these shares—he may even decide to offer some—until equation (1) of maximum satisfaction is extended to include $[\phi_e(q_e + d_e)]/p_e$.[7] Assuming that he demands some of these shares, the quantity demanded times price must enter into our budgetary equation (2) as $d_e p_e$ on the right side. In this way Walras attempts to handle an essentially dynamic concept with the aid of the tools of static theory. What the individual is purchasing when he buys a hypothetical perpetual annuity share is betterment. And in Walras' static system betterment is a commodity which can be bought and sold like any other commodity and for which an individual is presumed to have a desire or utility function like any other desire or utility function. The price of a unit of betterment (the reciprocal of the rate of capitalization) is thus determined in the same way as all other prices. Equations

[6] I have modified Walras' statement slightly by changing his π_p to p_p in order to avoid subsidiary issues not immediately relevant to the present argument.

[7] See first system of equations, (1), on p. 40, boldface addendum.

(1) and (2) as now amended are extremely interesting, for they make quite plain that there is no antithesis between Adam Smith's desire to better one's condition and Walras' desire to maximize satisfaction, for at any moment of time an individual can only undertake to better his condition within the limits of the resources he has at his disposal. From our modified (1) and (2) we obtain an additional demand function for equation (4), viz., $d_e = f_e(\ \ )$, after suitably amending the other demand and supply functions to include the price of the standard commodity among all the other prices. This demand function will behave like all other Walrasian demand functions and will become a supply function as the price continues to rise.

Walras describes a geometrical representation of this theory,[8]

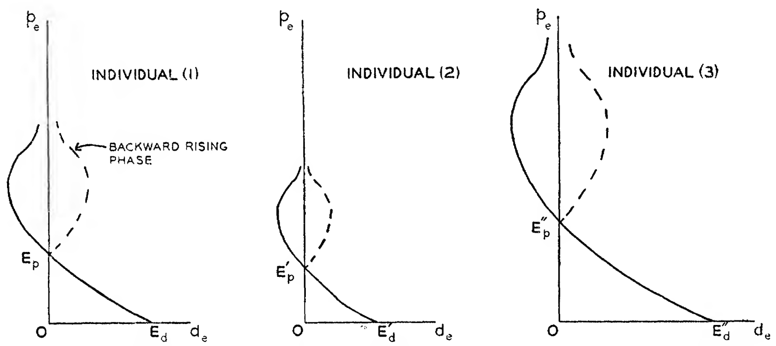


FIG. 1.—Individual demand and supply of perpetual annuity shares

which I am giving on Marshallian axes (Fig. 1) to make it look a little more familiar to readers brought up in the Cambridge tradition. It will be recognized at once that offer segments of these continuous curves are really dissaving functions. Hence the intersection of the aggregate social demand and social supply curves for these shares (Fig. 2) will only give us the price of the share $D'_e\,\beta = p'_e$ at which dissaving is

[8] *Eléments*, pp. 251–52. In Fig. 1, OE_d, OE'_d, and OE''_d represent the amounts of perpetual annuity shares taken by individuals (1), (2), and (3), respectively, when the price per share p_e is zero. These amounts decrease as p_e rises, and become zero for our individuals at E_p, E'_p, and E''_p, respectively. At still higher prices our individuals no longer demand, but offer, the shares. For example, as p_e rises above E_p, individual (1) first offers increasing amounts and then decreasing amounts. This is shown by the continuation of the $E_d\,E_p$ curve to the left of the vertical axis. The offer segment when pivoted around the vertical axis is brought (in the form of the broken-line curve) into the same quadrant as the demand curve. This is done for each individual. Figure 2 is, then, constructed by the horizontal summation of the individual demand curves, giving the social demand curve $E''_p\,\Sigma\,E_d$, and by the horizontal summation of the individual offer curves, giving the broken-line social supply curve $\alpha\beta\gamma$.

exactly equal to saving. The p'_e is, therefore, the price of the share in the stationary state. To obtain a demand curve for new shares, we must subtract horizontally the $\alpha\beta$ segment of the supply curve from the de-

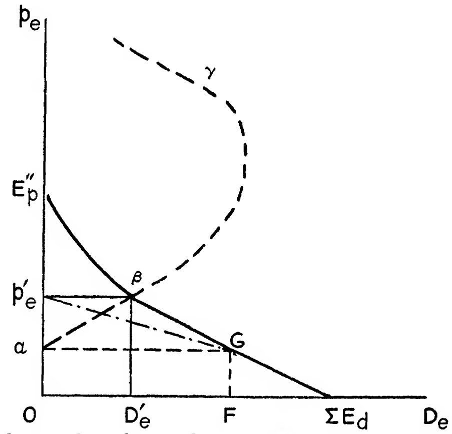


FIG. 2.—Social demand and supply of perpetual annuity shares

mand curve $E''_p \sum E_d$ which gives us the curve $p'_e G \sum E_d$. This is the curve of net demand for more perpetual annuity shares in our community, and the equation of that curve is

$$D_e = F_e(p_t, p_p, p_k, p_b, p_c, p_e) .$$

If we multiply both sides of this equation by p_e, we obtain the supply function of savings

$$D_e p_e = F_e(p_t, p_p, p_k, p_b, p_c, p_e) p_e ,$$

which may also be expressed — in view of the fact that $p_e = 1/i$ — as

$$D_e p_e = \mathfrak{F}_e(p_t, p_p, p_k, p_b, p_c, i) .$$

This we may add to our system of equations (I), at the same time that we add p_e to the independent variables of the supply and demand equations of systems (I) and (II).

Furthermore, we can translate geometrically our demand curve for perpetual annuity shares $p'_e G \sum E_d$ into a supply curve of savings as a function of the rate of net income $i = 1/p_e$. A decision to purchase, say, αG units of perpetual annuity shares is tantamount to a decision to save or set aside the equivalent of $O\alpha GF$ units of standard commodity when the rate of net income in $1/O\alpha$. Following this through for the reciprocals of all values of p_e from O to p'_e, and setting up abcissas proportional to the inscribed rectangles, we obtain the $i'\varepsilon T$ curve shown in Figure 3. It is obvious that in order to

find an equilibrium rate of net income that would correspond to an equilibrium rate of capital accumulation, we must have a demand for these net savings. Walras gives one, namely,

$$\frac{D_k \pi_k}{i} = D_k \frac{p_k}{i + \mu_k + \nu_k} = D_k P_k \ ,$$

which is obviously a decreasing function of i and which is represented by the curve $V_\varepsilon U$. Where the two curves intersect, $D_e p_e = D_k p_k$. Here D_k is defined as the quantity of capital goods *to be* manufactured;

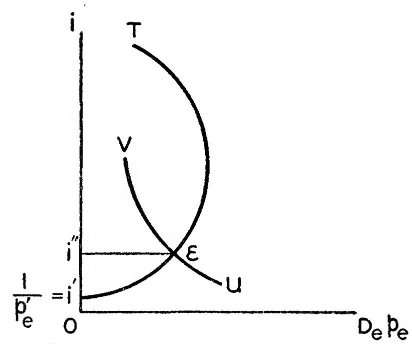


FIG. 3.—Supply and demand curves of net savings as a function of i, other things remaining equal.

and this quantity is determined at the point where the capitalized value of a unit of capital good is equal to its cost of production:

$$P_k = \frac{p_k}{i + \mu_k + \nu_k} = k_t p_t + k_p p_p + k_k p_k \ .$$

All this has the *appearance* of giving a satisfactory solution to the whole problem of capital accumulation. It gives us the negatively inclined demand curve $V\varepsilon U$ of our Figure 3. Also it enables us to add enough new equations to the system with which we started to make it mathematically determinate.

New Unknowns	New Equations
D_e	$D_e p_e = F_e(\ )$
D_k	$D_k P_k = D_e p_e$
P_k	$P_k = k_t p_t + k_p p_p + k_k p_k$
$P_e = \frac{1}{i}$	$P_k = \frac{p_k}{i + \mu_k + \nu_k}$

In spite of this apparent determinacy, it is not at all clear, to me at least, where the D_k came from. It seems to me to be an adventitious element which is neither directly nor indirectly related to the utility

functions which constitute, as it were, the primary motive for the whole system. Walras went to a great deal of trouble to make decisions to save an integral part of his equilibrium structure, but he gives no clue at all to the reasons for decisions to manufacture, or decisions to invest in, new capital goods. Whatever rationally determined demand for savings there was has already been taken care of in the dissaving function. The demand for still more savings on the part of the manufacturers of capital goods is not given a rational basis in Walras' system. If this view is correct, Walras' theory of capital accumulation remains indeterminate. And if I am right, the indeterminacy is due either to an inherent defect in Walras' static system or to an unfortunate attempt to squeeze an essentially dynamic phenomenon[9] into a static framework. I think the latter reason is more probable.

[9] Walras' theory of capital formation has every appearance of a static theory in that the mere decision of an individual to spend part of his income on some means of increasing his income has no effect at the moment of decision upon his *resources* and in that the mere decision to add to the existing stock of utilityless capital goods can neither arise from any change in *tastes* nor give rise immediately to any change in *technique*. If, however, these decisions are affected by *uncertain expectations* as to the future income to be derived from investments of savings in view, i.e., additional capital goods, then the phenomenon belongs properly to the realm of dynamics (see J. R. Hicks, *Value and Capital*, [London: Oxford University Press, 1939] chap. IX). Though Walras failed to incorporate the concept of expectations in his theoretical structure, he recognized its importance as an afterthought. At the very close of his discussion of capital accumulation we find the following passage, which I have translated from pp. 292 and 293 of the *Eléments*: "We must remember that the income of new capital goods is not known with the same degree of certainty as that of existing capital goods. It may prove to be larger or smaller than anticipated; in short, it is more problematical . . . . . Then, too, we must remember that the price of capital goods varies by reason, not only of past changes, but also of expected changes in gross income or in rates of depreciation and insurance; and that, so far as future changes are concerned, the expectations differ from individual to individual."

SAY'S LAW: A RESTATEMENT AND CRITICISM

OSCAR LANGE
University of Chicago and the Cowles Commission

SAY'S law is the proposition that there can be no excess of total supply of commodities (general oversupply) because the total supply of all commodities is *identically* equal to the total demand for all commodities. Under certain assumptions as to the nature of the demand for money this proposition appears as a simple corollary of the general theory of prices. Associated with it is the proposition that there cannot be such a shortage of total entrepreneurial receipts relative to total entrepreneurial cost as to cause losses throughout the whole economy (general overproduction). The present paper intends to investigate the relation of these propositions to each other and to study the implications of Say's law with regard to the problem of underemployment, to the general theory of prices, and to the theory of money.

2. Let us consider a closed system in which n commodities are exchanged, one of them—say, the nth commodity—functioning as medium of exchange as well as *numéraire,* i.e., as money. Denote by p_i the price of the ith commodity. We have $p_n \equiv 1$. Let $D_i = D_i(p_1, p_2, \cdots, p_{n-1})$ and $S_i = S_i(p_1, p_2, \cdots, p_{n-1})$ be the demand function and the supply function, respectively, of the ith commodity. The equilibrium prices are determined by the $n-1$ equations

$$D_i(p_1, p_2, \cdots, p_{n-1}) = S_i(p_1, p_2, \cdots, p_{n-1}). \qquad (i = 1, 2, \cdots, n-1) \quad (2.1)$$

The condition of stability of the equilibrium of the price system is expressed by the $(n-1)^2$ inequalities and equations[1]

$$\frac{dD_j}{dp_i} < \frac{dS_j}{dp_i} \quad \text{when } j = i. \qquad (i \text{ and } j = 1, 2, \cdots, n-1) \quad (2.2)$$
$$\frac{dD_j}{dp_i} = \frac{dS_j}{dp_i} \quad \text{when } j \neq i.$$

[1] Cf. J. R. Hicks, *Value and Capital* (London: Oxford University Press, 1939), pp. 66-67. This condition is sufficient. Hicks gives additional conditions for what he calls "perfect stability." The concept of perfect stability, however, refers to the *way* in which the stability of the system is maintained and need not occupy us here.

There are only $n - 1$ independent demand functions and $n - 1$ independent supply functions, the demand and the supply function for the commodity which functions as money being deducible from the other ones. We have

$$\sum_{i=1}^{n-1} p_i D_i \equiv S_n \tag{2.3}$$

and

$$\sum_{i=1}^{n-1} p_i S_i \equiv D_n. \tag{2.4}$$

Taking account of the last two relationships, we obtain the total demand (measured in money value) for all n commodities

$$\sum_{i=1}^{n} p_i D_i \equiv \sum_{i=1}^{n-1} p_i D_i + D_n \equiv S_n + D_n. \tag{2.5}$$

Similarly the total supply (measured in money value) of all n commodities is

$$\sum_{i=1}^{n} p_i S_i \equiv \sum_{i=1}^{n-1} p_i S_i + S_n \equiv D_n + S_n. \tag{2.6}$$

Therefore

$$\sum_{i=1}^{n} p_i D_i \equiv \sum_{i=1}^{n} p_i S_i , \tag{2.7}$$

i.e., total demand and total supply are identically equal.

I propose to call this identity *Walras' law* because Walras was the first to recognize its fundamental importance in the formulation of the mathematical theory of prices. It should be noted that Walras' law does not require that the demand and supply of each commodity, or of any of them, be in equilibrium. The identity of (2.7) holds independently of whether the equations (2.1) are satisfied or not.[2]

[2] Walras' law holds also in absence of a uniform medium of exchange, i.e., in a moneyless system. Let D_{ij} and S_{ij} be that part of the demand or supply, respectively, of the ith commodity for which the jth commodity is offered or demanded in exchange. Let, further, p_{ij} be the price of the ith commodity in terms of the jth. We have then

$$S_{ji} \equiv D_{ij} p_{ij} \tag{1}$$

and $(i \text{ and } j = 1, 2, \cdots, n)$

$$D_{ji} \equiv S_{ij} p_{ij} \tag{2}$$

Taking (arbitrarily) one of the commodities as *numéraire* and expressing all prices in terms of it, we have

$$p_{ij} = \frac{p_i}{p_j} ,$$

3. Now let us consider all commodities *exclusive* of money. To simplify the exposition, the term "commodity" will be henceforward understood to exclude money. Thus we shall oppose "commodities" to "money."

The total demand for commodities (exclusive of money) is $\sum_{i=1}^{n-1} p_i D_i$ and the total supply of commodities (exclusive of money) is $\sum_{i=1}^{n-1} p_i S_i$. From (2.3) and (2.4) it follows directly that

$$\sum_{i=1}^{n-1} p_i D_i = \sum_{i=1}^{n-1} p_i S_i \tag{3.1}$$

when and only when

$$D_n = S_n , \tag{3.2}$$

i.e., when the demand for money is equal to the supply of money.

But D_n and S_n are the demand for and the supply of money in a

---

where p_i and p_j are the price of the ith and of the jth commodity in terms of *numéraire*. Thus

$$p_j S_{ji} \equiv p_i D_{ij} \tag{3}$$

and

$$p_j D_{ji} \equiv p_i S_{ij} . \tag{4}$$

Total demand, expressed in *numéraire* units, for all n commodities is $\sum_{i=1}^{n} \sum_{j=1}^{n} p_i D_{ij}$ and total supply, similarly expressed, of all n commodities is $\sum_{i=1}^{n} \sum_{j=1}^{n} p_i S_{ij}$. On account of (4) we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} p_i S_{ij} \equiv \sum_{i=1}^{n} \sum_{j=1}^{n} p_j D_{ji} . \tag{5}$$

Because of symmetry of subscripts ($i = 1 , 2 , \cdots , n; j = 1 , 2 , \cdots , n$) we have also

$$\sum_{i=1}^{n} \sum_{j=1}^{n} p_j D_{ji} \equiv \sum_{i=1}^{n} \sum_{j=1}^{n} p_i D_{ij} ,$$

and substituting this in (5) we obtain

$$\sum_{i=1}^{n} \sum_{j=1}^{n} p_i D_{ij} \equiv \sum_{i=1}^{n} \sum_{j=1}^{n} p_i S_{ij} , \tag{6}$$

i.e., Walras' law. Walras' proof is somewhat different. He has proved the theorem that if demand equals supply for $n-1$ commodities it does so also for the nth commodity (cf. *Elements d'économie politique pure* ["édition définitive"; Paris and Lausanne, 1926], pp. 120–21). This implies that total demand equals identically total supply of all n commodities and is, therefore, equivalent to (6).

particular sense, namely, the money demanded in exchange for the commodities offered and the money offered in exchange for the commodities demanded. It is more convenient to express (3.2) in relation to the existing stock of money and to the demand for cash balances. A difference between the money demanded in exchange for commodities and the money offered in exchange for commodities implies a desire to change cash balances relative to the amount of money available. The desired change is equal to that difference. Let us denote by ΔM the total increase of cash balances (in excess of a possible increase of the quantity of money) desired by all individuals. We have then:[3]

$$D_n - S_n \equiv \Delta M . \tag{3.3}$$

Condition (3.2) may now be written in the form:

$$\Delta M = 0 , \tag{3.4}$$

i.e., there is no desire to change the total sum of cash balances relative to the quantity of money. This means that the total demand for cash balances is equal to the existing stock of money. Thus the necessary and sufficient condition that the total demand for commodities be equal to the total supply of commodities is that the total demand for cash balances be equal to the amount of money in existence. We may call the fulfilment of this condition *monetary equilibrium.*

The total demand for commodities is equal to total supply of commodities only in a state of monetary equilibrium.

4. Say's law makes a much stronger claim than either Walras' law or the equality of total demand for commodities and total supply of commodities under conditions of monetary equilibrium. It states that the total demand for commodities (exclusive of money) is *identically* equal to their total supply:

$$\sum_{i=1}^{n-1} p_i D_i \equiv \sum_{i=1}^{n-1} p_i S_i . \tag{4.1}$$

From (2.3) and (2.4) we see immediately that, in order that Say's law shall hold, it is necessary and sufficient that

$$D_n \equiv S_n , \tag{4.2}$$

which because of (3.3) can also be written

$$\Delta M \equiv 0 , \tag{4.3}$$

[3] D_n and S_n are, like all quantities demanded or supplied, measured as per unit or period of time. Consequently, ΔM is measured in the same way.

i.e., the total demand for cash balances must be *identically* equal to the amount of money in existence.

Thus Say's law implies a peculiar nature of the demand for money, namely, that the individuals in our system, taken together, are *always* satisfied with the existing amount of money and never wish to hold either more or less. There is never a desire to change the total cash balances otherwise than to adapt them to changes in the amount of money available. Under these circumstances purchases of commodities are never financed from cash balances nor do sales of commodities serve to increase cash balances.

This peculiar nature of the demand for money implied in Say's law was clearly understood by its original proponents. They assumed it explicitly by stating that money is only a medium of exchange and abstracting from its function as "store of value." In the *Traité d'économic politique*[4] Say states explicitly that, when there is an oversupply of certain commodities, the difficulty in selling them is only seemingly the lack of money to buy them. The lack of money, says Say, is but an expression of the lack of other commodities because the money to be offered for the purchase of the commodities of which there is an oversupply can be acquired only through the sale of other commodities. This view excludes the use of cash balances for financing purchases of commodities. The same view is also expressed by Ricardo: "Productions are always bought by productions, or by services; money is only the medium by which the exchange is effected."[5]

5. From its very first enunciation Say's law has been associated with the proposition that there can be no "universal glut" or "general overproduction" in the sense of all entrepreneurs suffering losses. As Ricardo puts it, in a sequel to the passage just quoted: "Too much of a particular commodity may be produced, of which there may be such a glut in the market as not to repay the capital expended on it; but this cannot be the case with respect to all commodities."[6] Total entrepreneurial receipts are thought of as being identically equal to total cost plus some measure of profit (to be discussed later), and a deficiency of receipts with respect to one commodity must, therefore, be accompanied by a surplus of receipts with respect to some other commodity (or commodities). "Overproduction" can be only "partial," each partial overproduction being accompanied by a partial underproduction somewhere else in the economic system. We shall investi-

[4] See pp. 347–48 of the *Traité* (Paris, 1861).
[5] *Principles of Political Economy and Taxation,* chap. xxi.
[6] *Ibid.*

gate the relation of this proposition to Say's law, with special regard to the nature of the "measure of profit" involved.

Let us distinguish between commodities bought by entrepreneurs and commodities sold by entrepreneurs. We shall call the first "factors" and the other "products." A commodity may be both a factor and a product, or it may be neither. Thus we get the following four classes of commodities: commodities which are only factors, commodities which are both factors and products, commodities which are only products, and, finally, commodities which are neither factors nor products. We shall call these four classes "primary factors," "intermediate products," "final products," and "direct services," respectively. To simplify the notation, let us denote the total demand for and the total supply (both measured in money value) of a class of commodities by D and S with a subscript indicating the class. Use the subscripts F, I, P, and C to denote primary factors, intermediate products, final products, and direct services, respectively. Let us further split up the demand for intermediate products into the demand for replacement of the intermediate products used up during the period in question (i.e., the period in terms of which the demand is measured) and the demand for *net* increase of the stock of intermediate products (new investment),[7] using the subscripts IR and IN to indicate the two types of demand for intermediate products. Finally, let us denote, as before, by D_n and S_n the demand and supply of money in exchange for commodities.

Since our classification is exhaustive, we have

$$\sum_{i=1}^{n} p_i D_i \equiv D_F + D_{IR} + D_{IN} + D_P + D_C + D_n$$

and

$$\sum_{i=1}^{n} p_i S_i \equiv S_F + S_I + S_P + S_C + S_n.$$

By Walras' law we have

$$(D_F + D_{IR}) + D_{IN} + D_P + D_C + \Delta M \equiv S_F + (S_I + S_P) + S_C, \tag{5.1}$$

where $\Delta M \equiv D_n - S_n$, as before. The part in parentheses on the left-

[7] Thus, if less than the amount of intermediate products used up during the period is replaced, the demand for *net* increase of stock (new investment) is negative. The demand for replacement represents what Keynes calls "user cost" and "supplementary cost" (see J. M. Keynes, *The General Theory of Employment* [New York: Harcourt Brace & Co., 1936], pp. 53 and 56). The actual demand for intermediate products is the demand for replacement plus the demand for *net* increase of stock.

hand side of this equation is the demand, measured in money value, of entrepreneurs for primary factors and for replacement of the intermediate products used up, i.e., the total cost entrepreneurs are ready to incur. The part in parentheses on the right-hand side is the supply of products, measured in money value, i.e., the total receipts planned by entrepreneurs. The difference between the two

$$\Pi \equiv (S_I + S_P) - (D_F + D_{IR}) \tag{5.2}$$

is the total profit entrepreneurs plan to receive.[8] This is the total profit implied in the entrepreneurs' decisions to offer $(S_I + S_P)$ worth of products and to use $(D_F + D_{IR})$ worth of factors.[9] We shall call it *planned* total profit.

Taking into account (5.2), we can re-write (5.1) in the form

$$(\Pi - D_{IN}) - (D_P - S_F) \equiv \Delta M - \Delta C\,, \tag{5.3}$$

where

$$\Delta C \equiv S_C - D_C\,.$$

Each of the terms in this identity, except ΔM, represents an independent set of decisions. The terms in the first parentheses represent entrepreneurial decisions, and the terms in the second parentheses represent decisions to buy from entrepreneurs and to sell to entrepreneurs. We shall call these decisions the *capitalistic sphere* of decisions. On the right-hand side the term ΔC represents decisions to sell and to buy direct services. As the offers to sell and to buy direct services are not directed to entrepreneurs, we shall call these decisions the *noncapitalistic sphere* of decisions. On account of (5.2)

$$\Pi - D_{IN} \equiv (S_I + S_P) - (D_F + D_{IR} + D_{IN})\,,$$

which is the difference between the stream of money demanded and the stream of money offered by entrepreneurs. It is the *net* stream of money demanded by entrepreneurs. The expression $(D_P - S_F)$ is the difference between the stream of money offered to entrepreneurs and the stream of money demanded from entrepreneurs. It is the *net* stream of money offered to entrepreneurs. On the right-hand side

[8] It is assumed here that entrepreneurs supply exactly the quantities designed, i.e., the quantities indicated by the supply functions. If entrepreneurs' supply is different from what was designed by them (as, e.g., in case of fluctuating crops), Π differs from the profit entrepreneurs plan to receive by the difference between the actual supply and the supply originally designed.

[9] S_I and S_P are expressions of the type $\Sigma p_i S_i$, the summation extending over all intermediate products and all final products, respectively. D_F and D_{IR} are expressions of the type $\Sigma p_i D_i$, the summation extending over the respective class of commodities. The S_i's and D_i's are functions of prices. The prices are taken as (arbitrarily) given.

$\Delta C \equiv S_c - D_c$ is the difference between the stream of money demanded and the stream of money offered in exchange for direct services, or the demand for increase of cash balances arising in the noncapitalistic sphere of decisions. As ΔM is the total demand for increase of cash balances (relative to the quantity of money available), $\Delta M - \Delta C$ is the demand for increase of cash balances (relative to the quantity of money) arising in the capitalistic sphere of decisions. We shall say that there is monetary equilibrium in the capitalistic sphere of decisions when $\Delta M - \Delta C = 0$.

When $(D_P - S_F) = (\Pi - D_{IN})$, the *net* stream of money offered to entrepreneurs is equal to the *net* stream of money demanded by them, and entrepreneurs can realize their planned total profit and their demand for new investment. However, when $(D_P - S_F) < (\Pi - D_{IN})$, the net stream of money offered to entrepreneurs is less than the net stream of money demanded by them. Given their demand for new investment, entrepreneurs cannot realize their planned total profit. They must either accept a smaller total profit than planned or, instead, increase D_{IN}, i.e., their demand for new investment. Demand for new investment is an offer of entrepreneurs to buy from themselves. Therefore, an increase in the demand for new investment diminishes the net stream of money demanded by entrepreneurs. Finally, when $(D_P - S_F) > (\Pi - D_{IN})$, the net stream of money offered to entrepreneurs is greater than the net stream of money demanded by them. Entrepreneurs can either realize a greater total profit than planned or, alternatively diminish their new investment.

Thus, given the entrepreneurs' demand for new investment, D_{IN}, a profit less than, equal to, or greater than planned can be realized according as to whether $(D_P - S_F) \lesseqgtr (\Pi - D_{IN})$ or, because of (5.3), according as to whether $\Delta M - \Delta C \gtreqless 0$. It should be noticed that the condition which permits entrepreneurs to realize exactly their planned total profit and their demand for new investment is not equivalent to monetary equilibrium for the whole system ($\Delta M = 0$) but to monetary equilibrium in the capitalistic sphere of decisions (i.e., $\Delta M - \Delta C = 0$). However, in a purely capitalistic system (i.e., in a system in which there are not direct services) ΔC disappears and $\Delta M = 0$ is the condition equivalent with the equality between the net stream of money offered to entrepreneurs and the net stream of money demanded by entrepreneurs.

Under Say's law $\Delta M \equiv 0$, and if the economic system is purely capitalistic we have

$$(D_P - S_F) \equiv (\Pi - D_{IN}). \tag{5.4}$$

The net stream of money offered to entrepreneurs is *always* equal to the net stream of money demanded by them. Whatever the total profit and new investment planned by entrepreneurs, the net stream of money offered to them is always such as to enable them to realize their planned profit and new investment, irrespectively of whether there is equilibrium of demand and supply of each separate commodity. Thus total entrepreneurial receipts are, under Say's law, identically equal to total cost plus planned total profit.[10] Consequently, an impossibility of realizing planned profit in one part of the system must be compensated by a possibility of realizing more than planned profit in some other part of the system. It is in this sense that "overproduction" can be only "partial."

This holds, however, only for a purely capitalistic system. If direct services are present, the condition $\Delta M \equiv 0$ does not suffice to make that total entrepreneurial receipts equal to total cost plus planned total profit. It is necessary, in addition, that $\Delta C = 0$, i.e. that the market for direct services be in equilibrium. Disequilibrium in the market for direct services may cause a discrepancy between the net stream of money offered to entrepreneurs and the net stream of money demanded by entrepreneurs. Thus Say's law implies the impossibility of a "universal glut" only for a purely capitalistic system.

6. Say's law, however, does *not* imply that the total demand and the total supply of products are identically equal. Neither does it imply an identity of the total demand and the total supply of primary factors and direct services.

We can re-write (5.1) in the form

$$\begin{aligned}(S_F + S_C) &- (D_F + D_C) \\ &\equiv (D_{IR} + D_{IN} + D_P) - (S_I + S_P) + \Delta M .\end{aligned} \tag{6.1}$$

The left-hand side is the excess supply of factors and direct services (measured in money value). On the right-hand side the two parts in parentheses give the excess demand for (intermediate and final) products (also measured in money value).[11] The identity (6.1) shows that primary factors and direct services are offered in exchange for products and money, while products and money are offered in ex-

[10] This is subject to the qualification stated in n. 8. It should also be noticed that Say's law implies nothing as to the level of planned total profit. Planned total profit may be even negative, as, e.g., when there is large fixed capital equipment.

[11] By "excess supply" we mean the excess of supply over demand; by "excess demand," the excess of demand over supply.

change for primary factors and direct services. Under conditions of monetary equilibrium $\Delta M = 0$, and an excess supply of factors implies an excess demand of equal size for products, and vice versa. This is clear, because, when monetary equilibrium exists, primary factors and direct services are offered in exchange for products alone, and products alone are offered in exchange for primary factors and direct services. In this case an excess supply of the one means an excess demand for the other.

Under Say's law $\Delta M \equiv 0$, and we obtain

$$(S_F + S_C) - (D_F + D_C) \equiv (D_{IR} + D_{IN} + D_P) - (S_I + S_P). \tag{6.2}$$

The two sides of this identity need not be zero. The total supply of primary factors and direct services may, therefore, differ from the total demand for primary factors and direct services. Similarly, the total demand for products may differ from the total supply of products. Neither of the two discrepancies is precluded by Say's law. But (6.2) shows that under Say's law an excess supply of primary factors and direct services *always* implies an excess demand of equal amount for products, and vice versa. This tends directly to restore equilibrium. An excess demand for products causes product prices to rise. This stimulates a decrease of the demand and an increase of the supply of products. A decrease of demand for products is, however, equivalent to a decrease of supply of primary factors and direct services; and an increase of supply of products is equivalent to an increase of demand for primary factors and direct services. Thus equilibrium is restored simultaneously between demand and supply of products and between demand and supply of primary factors and direct services. In a similar way equilibrium is restored directly in case of an excess supply of products.

When Say's law does not hold, an excess supply of primary factors and direct services need not be associated with an excess demand for factors. From (6.1) we see immediately that an excess supply of products may coexist with an excess supply of factors and direct services, and vice versa, as long as $\Delta M > 0$. Such a coexistence happens when

$$\Delta M > (S_I + S_P) - (D_{IR} + D_{IN} + D_P) > 0$$

and

$$\Delta M > (S_F + S_C) - (D_F + D_C) > 0\,,$$

i.e., when there is a desire to increase cash balances (relative to the

quantity of money available) by more than the excess supply of products, and also by more than the excess supply of factors and direct services. In this case there is no direct tendency to restore equilibrium through the simple mechanism of exchange between primary factors and direct services on one side and products on the other side. Equilibrium can be restored only through abatement of the desire to increase cash balances relative to the quantity of money (i.e., through ΔM's becoming again equal to zero). This will happen only if the fall in prices resulting from the excess supply tends to make $\Delta M = 0$. We may say that in such cases the conditions for a stable monetary equilibrium are satisfied. Otherwise there is no tendency to reach an equilibrium, and the general stability conditions (2.2) are not satisfied. However, the satisfaction of *all* the stability conditions (2.2) is not implied in Say's law. Say's law implies only that enough of the stability conditions of the system hold, to assure the existence of a stable equilibrium for two broad classes of commodities, namely, the class of products and the class of primary factors and direct services.

This discussion takes us back to the original controversy between Malthus and Ricardo. In his *Principles of Political Economy* Malthus stated: "If commodities were only to be compared and exchanged with each other, then indeed it would be true that, if they were all increased in their proper proportion, to any extent, they would continue to bear among themselves the same relative value."[12] But this is not a correct account, Malthus argued, because "it is by no means true, as a matter of fact that commodities are always exchanged for commodities. The great mass of commodities is exchanged directly for labour, either productive or unproductive; and it is quite obvious that this mass of commodities, compared with the labour with which it is to be exchanged, may fall in value from a glut just as any one commodity falls in value from an excess of supply, compared either with labour or money."[13] Malthus means here by the term "commodity" (as distinguished from "labor") what is designated in this paper as "product." Remembering this, and substituting "primary factors and direct services" for the more restricted concept of "labor" used by Malthus, the statement quoted expresses the relationship stated in (6.1) when $\Delta M = 0$. It states that products are exchanged not only for products but that "the great mass" of products is exchanged for primary factors and direct services. In consequence there may be a general excess supply of products which leads to a decline of product

[12] London, 1820, p. 355.

[13] *Ibid.*, pp. 353-54. Cf. also the note on pp. 317-18 of the second edition (London, 1836).

prices as compared with prices of primary factors and direct services.

This statement is quite correct, because an excess supply of products implies under these conditions an excess demand for factors and direct services. Malthus, however, thought that by pointing out this correct relationship he had proved *eo ipso* the possibility of a "general overproduction" (as defined above on p. 53). The effect of an excess supply of products would be, according to Malthus, that "commodities [i.e., products] would necessarily fall in value, compared with labour [primary factors and direct services], so as to lower profits almost to nothing, and to check for a time further production. But this is precisely what is meant by the term glut, which, in this case, is evidently general not partial."[14] It is true, as we have seen, that the prices of products would fall relatively to the prices of primary factors and direct services. But, as is clearly seen from (5.3), no "general glut" in the sense of realized total profit falling below planned total profit follows, unless $\Delta M - \Delta C > 0$.

Ricardo's [15] answer to Malthus was: "It is quite true that commodities may exist in such abundance, compared with labour, as to make their value so [to] fall, [estimated] in labour, as not to afford any inducement to their further production. In that case labour will demand a great quantity of commodities."[16] Keeping in mind our preceding observation concerning terminology, this passage simply states that an excess supply of products is accompanied by an excess demand for primary factors and direct services, which causes a rise

[14] *Op. cit.* (1820), p. 354.

[15] Say's reply to the same point made by Malthus was merely a terminological evasion: "Commodities, you say, are not only exchanged for commodities: they are also exchanged for labor. If this labor be a produce that some persons sell, that others buy, and that the latter consume, it will cost me very little to call it a *commodity*, and it will cost you very little more to assimilate other commodities to it, for they are also produce. Then comprising both under the generic name of *Produce*, you may perhaps admit that produce is bought only with produce" (cf. *Letters to Thomas Malthus on Political Economy and Stagnation of Commerce* [London, 1821 (reprinted in 1936 by Harding Ltd.)], Letter 1, pp. 21-22). In this translation the word "not" is omitted in the first sentence. This is obviously a misprint (cf. the French original in *Oevres diverses de J. B. Say, Petit volume*, ed. Guillaumin [Paris, 1848], p. 456). The word "produce" obviously means here commodities in general, i.e., primary factors and direct services as well as products.

At a later stage of the controversy Say fell only deeper into his terminological tautologies and defined "produce as a product the receipts for which cover its costs" (see the letter to Malthus of July, 1827, in *Oevres diverses*, p. 513, and *Cours complet d'économie politique pratique* [2d ed., 1840], I, 347–48); cf. also E. von Bergmann, *Geschichte der nationalökonomischen Krisentheorien* [Stuttgart: Kohlkammer, 1895], pp. 74-76). Thus the proposition that total cost of "produce" cannot exceed total receipts of entrepreneurs became with Say a mere tautology.

[16] *Notes on Malthus' "Principles of Political Economy,"* ed. Jacob H. Hollander and T. E. Gregory (Baltimore: Johns Hopkins University Press, 1928), p. 163.

in their prices. This is correct under the assumptions of Say's law. Malthus, however, denied that an excess supply of products must be associated with an excess demand for primary factors and direct services. He maintained that there will be also an excess supply of primary factors and direct services, i.e., unemployment.[17] This requires, as we have seen, absence of monetary equilibrium, namely, a desire to increase cash balances (relative to the quantity of money available) by more than the excess supply of products, and more than the excess supply of factors and direct services. In such case a "universal glut" may occur, indeed, provided that some of the demand for increase of cash balances arises in the capitalistic sphere of decisions (i.e., that $\Delta M - \Delta C > 0$).

Malthus had clearly something like this in his mind as is shown by the following statement in a footnote:

> Theoretical writers in Political Economy, from the fear of appearing to attach too much importance to money, have perhaps been too apt to throw it out of their consideration in their reasonings. It is an abstract truth that we want commodities, not money. But, in reality, no commodity for which it is possible to sell our goods at once, can be an adequate substitute for a circulating medium, and enable us in the same manner to provide for children to purchase an estate, or to command labor and provisions a year or two hence. A circulating medium is absolutely necessary to any considerable saving; and even the manufacturer would get on but slowly, if he were obliged to accumulate in kind all the wages of his workmen. We cannot therefore be surprised at his wanting money rather than commodities.[18]

But the fact that he had relegated to a footnote this crucial monetary consideration made his argument unconvincing to Ricardo, who argued throughout on the basis of the assumption that money is only a medium of exchange (i.e., $\Delta M \equiv 0$). Because Malthus had failed to make fully explicit his assumption concerning the demand for money, the discussion between him and Ricardo was carried on at cross-purposes.

7. Now let us study the implication of Say's law for the theory of

[17] *Ibid.*, pp. 361-62. The excess supply of primary factors and direct services is, however, not the same as "involuntary unemployment" in the Keynesian sense. "Involuntary unemployment," as defined in the Keynesian theory, is not an excess of supply of labor but *an equilibrium position* obtained by intersection of a demand and of a supply curve, the supply curve, however, being infinitely elastic with respect to money wages over a wide range, and the point of intersection being to the left of the region where the elasticity of supply of labor with respect to money wages becomes finite. Thus the left-hand side of (6.1) is always zero in the Keynesian theory. The different levels of employment refer to different levels of the demand and supply of labor (cf. Keynes, *op. cit.*, p. 15; and also my article, "The Rate of Interest and the Optimum Propensity to Consume," *Economica*, February, 1938, p. 31).

[18] *Principles of Political Economy* (1820), pp. 361–62.

prices. In the general case the $n-1$ equilibrium prices are determined by the $n-1$ equations (2.1) which express, for each commodity, the equality of demand and supply. If the stability conditions (2.2) are satisfied, the actual prices tend toward the equilibrium prices given by (2.1). However, when Say's law holds, the number of independent equations is reduced by one. According to Say's law

$$\sum_{i=1}^{n-2} p_i D_i + p_{n-1} D_{n-1} \equiv \sum_{i=1}^{n-2} p_i S_i + p_{n-1} S_{n-1} \tag{7.1}$$

(where the commodity $n-1$ is chosen arbitrarily). This expression shows that, if $D_i = S_i$ for the $n-2$ first commodities, we have necessarily $D_{n-1} = S_{n-1}$. We also have $D_n \equiv S_n$ by Say's law. The number of independent equations is only $n-2$, while the number of equilibrium prices to be determined is $n-1$. Thus, when Say's law holds, the equilibrium prices are indeterminate. The equations (2.1) determine in this case $n-2$ prices as functions of the price of the commodity $n-1$ (which is chosen arbitrarily), i.e., $p_i = f_i(p_{n-1})$. $(i = 1, 2, \cdots, n-2)$.[19]

[19] Denoting $F_i(p_1, p_2, \cdots, p_{n-1}) = D_i - S_i$, the $n-2$ independent equilibrium equations of the set (2.1) can be written in the form:

$$F_i(p_1, p_2, \cdots, p_{n-1}) = 0. \qquad (i = 1, 2, \cdots, n-2) \tag{1}$$

A solution with respect to $p_1, p_2, \cdots, p_{n-2}$ exists if

$$\frac{\partial(F_1, F_2, \cdots, F_{n-2})}{\partial(p_1, p_2, \cdots, p_{n-2})} \neq 0,$$

$p_1, p_2, \cdots, p_{n-2}$ being then functions of p_{n-1}. The Jacobian has the required property when the stability conditions are satisfied. The stability conditions of the system (1) are (see eq. [2.2] above and also Hicks, *op. cit.*, p. 315):

$$\begin{aligned}
\frac{dF_1}{dp_i} &= \sum_{r=1}^{n-2} \frac{\partial F_1}{\partial p_r} \frac{dp_r}{dp_i} = 0, \\
&- - - - - - - - \\
\frac{dF_i}{dp_i} &= \sum_{r=1}^{n-2} \frac{\partial F_i}{\partial p_r} \frac{dp_r}{dp_i} < 0, \qquad (i = 1, 2, \cdots, n-2) \\
&- - - - - - - - \\
\frac{dF_{n-2}}{dpi} &= \sum_{r=1}^{n-2} \frac{\partial F_{n-2}}{\partial p_r} \frac{dp_r}{dp_i} = 0.
\end{aligned} \tag{2}$$

Solving the system (2) with respect to dF_i/dp_i, we obtain

$$\frac{dF_i}{dp_i} = \frac{\partial(F_1, F_2, \cdots, F_{n-2})}{\partial(p_1, p_2, \cdots, p_{n-2})} \div \frac{\partial(F_1, \cdots, F_{i-1}, F_{i+1}, \cdots, F_{n-2})}{\partial(p_1, \cdots, p_{i-1}, p_{i+1}, \cdots, p_{n-2})}.$$
$$(i = 1, 2, \cdots, n-2)$$

Since this must be negative the numerator must be different from zero.

This indeterminateness of equilibrium prices which results from the acceptance of Say's law is, however, reduced considerably by taking account of the consequences of the peculiar nature of the demand for money implied in Say's law. Say's law precludes substitution between money and commodities because it implies that purchases of commodities cannot be financed from cash balances and that cash balances cannot be increased out of the receipts from the sale of commodities.[20] This has an important consequence for the structure of the demand and supply functions of commodities. These functions are derived from the theory of substitution. According to the principles of the theory of substitution, a change of the ratios of the prices of the different commodities leads, as a rule,[21] to a substitution of commodities the prices of which are relatively lowered for commodities the prices of which are relatively increased.[22] A proportional change of the prices of all commodities, i.e., of $p_1, p_2, \cdots, p_{n-1}$, implies a change of the exchange ratio of commodities for money (the price of money $p_n \equiv 1$ by definition). In the general case this would result in a substitution of money for commodities or vice versa. Say's law, however, precludes such a substitution. Thus, in the case in which Say's law holds, a proportional change of the prices of all commodities cannot affect the demand and supply of commodities relative to the demand and supply of money. But a proportional change of all prices does not induce a substitution between different commodities either. Therefore, the demand and supply functions of commodities are, when Say's law holds, homogeneous of zero degree, i.e., a proportional change of *all* prices does not affect the quantities demanded or offered. These quantities depend merely on the *relative prices,* i.e., on the ratios of the prices

$$\frac{p_1}{p_{n-1}}, \frac{p_2}{p_{n-1}}, \cdots, \frac{p_{n-2}}{p_{n-1}},$$

where the commodity $n - 1$ is chosen arbitrarily.

Denoting the relative prices by $\pi_i = \dfrac{p_i}{p_{n-1}}$ $(i = 1, 2, \cdots, n-2)$,

the equations expressing, for each commodity, the equilibrium of demand and supply can be written

[20] Cf. p. 53 above.
[21] This rule may be counteracted by complementarity.
[22] This holds also for the substitution between factors and products if factors are considered as negative products (cf. Hicks, *op. cit.*, p. 93 and pp. 319-22).

$$D_i(\pi_1, \pi_2, \cdots, \pi_{n-2}) \\ = S_i(\pi_1, \pi_2, \cdots, \pi_{n-2}) . \quad (i = 1, 2, \cdots, n-2) \tag{7.2}$$

They take the place of the $n-2$ independent equations among the equilibrium equations (2.1), and the equilibrium values of the $n-2$ relative prices are determinate.

Thus it is possible to determine the equilibrium values of the relative prices, i.e., of the ratios of the money prices, of commodities. The money prices, however, remain indeterminate.

8. Under Say's law the relative prices of commodities are found to be independent of the quantity of money in the system. Money is "neutral,"[23] or, to use the phrase of the classical economists, it is merely a "veil" which can be removed and relative prices can be studied as if the system were based on barter. Indeed, by precluding the substitution of money for commodities or vice versa, Say's law constructs a system which is equivalent to a barter economy. Money in such a system is merely a worthless medium of exchange and a standard of value.

The money prices of commodities are indeterminate in a system in which Say's law is satisfied. In order to determine them, we need to know the price p_{n-1} (the commodity $n-1$ being chosen arbitrarily). If this is known, the money prices can be obtained from the relative prices by the relation $p_i = p_{n-1}\pi_i$ $(i = 1, 2, \cdots, n-2)$. The price p_{n-1}, however, cannot be obtained under Say's law because we have only $n-2$ independent equations of equilibrium of demand and supply. This has led the traditional theory of money to determine the price p_{n-1} from a supplementary equation—the "equation of exchange"—introduced into the system. This equation can be written

$$k \sum_{i=1}^{n-1} p_i S_i = M , \tag{8.1}$$

where k is a constant, expressing the proportion of the total supply of commodities, measured in money value, which people want to hold in cash balances, i.e., k is the reciprocal of the velocity of circulation of money. M is the quantity of money.

Since $p_i = p_{n-1}\pi_i$ (for $i = 1, 2, \cdots, n-2$; for $i = n-1$ we put $\pi_i \equiv 1$ by definition), the equation (8.1) transforms into

$$k p_{n-1} \sum_{i=1}^{n-1} \pi_i S_i = M . \tag{8.2}$$

[23] Cf. the definition of "neutral" money in J. Koopmans, *Das neutrale Geld* ("Beiträge zur Geldtheorie"), ed. F. A. Hayek (Wien: Springer, 1933), p. 228.

The equilibrium values of the relative prices π_i are determined by the equations (7.2) and the equilibrium quantities of commodities supplied S_i $(i = 1, 2, \cdots, n-1)$ are obtained by substituting the π_i's into the supply functions.[24] The π_i's and S_i's thus given, p_{n-1} is determined from (8.2).

This is the procedure of the traditional theory of money. It implies a division of the theory of prices in two separate parts: (1) the determination of relative prices and (2) the determination of a multiplier (the "price level") by a monetary equation distinct from the system of equilibrium equations. It results in money being "neutral."[25]

This procedure, however, is self-contradictory. Equation (8.2) is not compatible with Say's law. The left-hand side of this equation is the total demand for cash balances, and the right-hand side is the existing stock of money. The difference is the desired change in cash balances (relative to the quantity of money). We have thus:[26]

$$kp_{n-1} \sum_{i=1}^{n-1} \pi_i S_i - M = \Delta M \ . \tag{8.3}$$

If p_{n-1} has a value which does not satisfy (8.2), there is a discrepancy between the amount of money people want to hold and the quantity of money in existence. This implies a discrepancy between the total demand and the total supply of commodities (see eq. [3.3]). Say's law, however, requires that $\Delta M \equiv 0$ (see eq. [3.4] above). In this case we obtain

$$kp_{n-1} \sum_{i=1}^{n-1} \pi_i S_i \equiv M \ , \tag{8.4}$$

i.e., an identity, which holds for *any* value of p_{n-1} and, therefore, cannot serve to determine p_{n-1}. But k cannot be constant and must be indeterminate to adjust itself to any value of p_{n-1} so that the identity be satisfied. Say's law implies an indeterminate velocity of circulation $(1/k)$ and the money prices are indeterminate.

Thus the traditional procedure of the theory of money involves a contradiction. Either Say's law is assumed and money prices are

[24] S_{n-1} is also obtained because we have $n-1$ supply functions of commodities, although there are only $n-2$ independent equilibrium equations.

[25] Except for "frictions" and time lags, which is in this case the only way through which money can influence the relative prices of commodities.

[26] In order that the ΔM here be the same as the ΔM in (3.3), it is necessary that the unit or period of time over which the change is contemplated be the same as the unit or period of time per which the quantities demanded and supplied are measured. Cf. n. 3 above.

indeterminate or money prices are made determinate—but then Say's law and hence the "neutrality" of money must be abandoned. Say's law precludes any monetary theory.

9. We have seen that Say's law precludes any monetary theory. The theory of money must, therefore, start with a rejection of Say's law. Instead of assuming that total demand and total supply of commodities are identically equal or, what is equivalent, that the total demand for cash balances is identically equal to the amount of money available, these identities have to be replaced by genuine equations. The objective of the theory of money is then to study the conditions under which equilibrium of total demand and of total supply of commodities (or, instead, equilibrium of total demand for cash balances and the quantity of money available) obtains and the processes by which such equilibrium is attained.

This objective was expressed very clearly by Wicksell: "Any theory of money worthy of the name must be able to show how and why the monetary or pecuniary demand for goods exceeds or falls short of the supply of goods in given conditions."[27] Wicksell also observed the difficulty of reconciling this with Say's law. He finally appeased his conscience by stating that total demand and total supply must be equal only "ultimately" but may differ "in the first place."[28] With this observation Wicksell, and with him all monetary theorists, gave up Say's law by substituting for the identity an equation which holds only in equilibrium. The statement that total demand and total supply *tend* to be equal "ultimately" is nothing but an assertion that the stability conditions for the system are satisfied. If the stability conditions (2.2) are satisfied, any disturbance of equilibrium will make the demand and supply of each commodity tend toward equality again; and, since this happens for each commodity in the system, it also implies that total demand and total supply of commodities tend toward equality. But this tendency toward equilibrium, implied in the stability conditions, should not be confused with Say's law.

Since the homogeneity of the demand and supply functions of commodities disappears when Say's law is abandoned, we find that the theory of money cannot be separated from the theory of relative prices. The very basis of monetary theory proves to be incompatible with "neutrality" of money. The money prices of all commodities have to be determined directly from the general system of equilibrium equations (2.1).

[27] *Lectures on Political Economy,* II (London: Routledge & Sons, 1935), 159-60.

[28] *Ibid.,* p. 159.

10. The above implications of Say's law for the theory of prices and the theory of money hold also with regard to a dynamic theory of prices which is based on consideration of substitution of goods at different moments of time as well as of substitution of different goods at a given moment of time.

For simplicity let us divide the span of time under consideration in $m+1$ small and equal intervals indicated by the indices $0, 1, 2, \cdots, m$, where the index 0 refers to the "present" interval, the other indices referring to "future" intervals. Denote, further, the price of the ith commodity expected in the tth interval by p_{it} and let it be understood that $p_{io} (i = 1, 2, \cdots, n-1)$ are the prices actually obtaining in the "present" interval. We shall call the latter the "current prices." Let r_t be the rate of interest (per interval) on loans of a duration of t intervals. The discounted value of the expected price p_{it} is $q_{it} = p_{it}/(1+r_t)^t$. This definition yields $q_{io} = p_{io}$ for $i = 1, 2, \cdots, n-1$. Current demand and supply of a commodity, i.e., demand and supply in the "present" interval, is a function of all current prices as well as of the discounted values of all expected future prices[29]

$$D_{io} = D_{io}(q_{10}, q_{20}, \cdots, q_{n-1,0}; q_{11}\ q_{21}, \cdots, q_{n-1,1}; \cdots; \\ q_{1m}, q_{2m}, \cdots, q_{n-1,m}) \qquad (i = 1, 2, \cdots, n-1)$$

and

$$S_{io} = S_{io}(q_{10}, q_{20}, \cdots, q_{n-1,0}; q_{11}, q_{21}, \cdots, q_{n-1,1}; \cdots; \\ q_{1m}, q_{2m}, \cdots, q_{n-1,m}).$$

The equations of equilibrium are

$$D_{io}(p_{10}, p_{20}, \cdots, p_{n-1,0}; q_{11}, q_{21}, \cdots, q_{n-1,1}; \cdots; q_{1m}, q_{2m}, \cdots, q_{n-1,m}) = \\ S_{io}(p_{10}, p_{20}, \cdots, p_{n-1,0}; q_{11}, q_{21}, \cdots, q_{n-1,1}; \cdots; q_{1m}, q_{2m}, \cdots, q_{n-1,m}). \\ (i = 1, 2, \cdots, n-1) \qquad (10.1)$$

They determine the equilibrium values of the $n-1$ current prices $p_{io} (i = 1, 2, \cdots, n-1)$, as functions of the discounted values of the expected future prices. The latter may be regarded as functions of the current prices

$$q_{it} = f_{it}(p_{10}, p_{20}, \cdots, p_{n-1,0}). \\ (i = 1, 2, \cdots, n-1; t = 1, 2, \cdots, m) \qquad (10.2)$$

We shall call these functions the "expectation functions" and their

[29] Cf. Gerhard Tintner, "The Theoretical Derivation of Dynamic Demand Curves," *Econometrica,* October, 1938; and Hicks, *op. cit.,* chap xviii.

partial elasticities the "elasticities of expectation."[30] Thus, together with the expectation functions, which are $(n-1)m$ in number, the equations (10.1) determine the equilibrium values of the current prices.

When Say's law holds, we have, as before, only $n-2$ independent equations among the equations (10.1), and the demand and supply functions are homogeneous of zero degree because Say's law precludes substitution between money and commodities. In the dynamic theory of prices it is, however, *all* the money prices q_{it}, the discounted values of the expected future prices as well as the current prices, a proportional change of which does not affect the quantities demanded and offered. The demand and supply functions depend then only on the relative prices, i.e., on the ratios of the q_{it}'s. This, however, does not suffice to make the relative prices determinate because of the expectation functions (10.2). In order that the relative prices be determinate, the expectation functions, too, must involve only relative prices and not money prices. Thus the expectation functions must be homogeneous of the first degree, i.e., a proportional change of all current prices must change the discounted values of the expected future prices in the same proportion. In this case a proportional change of all current prices leaves the quantities demanded and supplied unaffected. The demand and supply functions of commodities depend now only on the ratios of the current prices, and the relative prices are determined by the $n-2$ independent equations of the system (10.1) and by the expectation functions (10.2). The money prices, however, remain indeterminate.

In the dynamic theory of prices Say's law implies thus, in addition to homogeneity of the demand and supply functions of commodities, homogeneous expectation functions. This additional assumption makes Say's law much more unrealistic in the context of a dynamic theory of prices than it is in the context of static theory. Both in static and in dynamic theory Say's law leaves money prices indeterminate.

[30] The latter term was introduced by Hicks (*op. cit.*, p. 205).

Studies in Mathematical Economics and Econometrics in Memory of Henry Schultz, ed. O. Lange, F. McIntyre, and T. Yntema, Chicago: University of Chicago Press, 1942

ON THE INTERPRETATION OF THE FUNDAMENTAL EQUATION OF VALUE THEORY

JACOB L. MOSAK
University of Chicago and the Cowles Commission

IN THIS note I desire to call attention to an interesting difference between Hicks and Slutsky in the interpretation of the fundamental equation of value theory. As was first shown by Slutsky in 1915,[1] the change in an individual's demand for a consumer good brought about by a change in price is given by the equation[2]

$$\frac{\partial X_s}{\partial p_r} = - X_r \frac{\partial X_s}{\partial M} + \mu \frac{U_{rs}}{U} \tag{1}$$

$$= - X_r \frac{\partial X_s}{\partial M} + X_{rs} \, .$$

$$(r \text{ and } s = 1, 2, \cdots, n)$$

Hicks interprets this equation as follows:

A fall in the price of a commodity . . . affect[s] the demand for that commodity in two different ways. On the one hand, it makes the consumer better off,

[1] Eugenio Slutsky, "Sulla teoria del bilancio del consumatore," *Giornale degli economisti*, LI (1915), 1–26.

[2] The notation is that used by J. R. Hicks in his *Value and Capital* (London: Oxford University Press, 1939). In this equation

X_s	represents a consumer good $(s = 1, 2, \ldots, n)$	(1)
p_r	is the price of X_r $\quad r = 1, 2, \ldots, n)$	(2)
M	$= \Sigma p_r X_r$ is the individual's income	(3)
u	$= f(X_1, X_2, \ldots, X_n)$ is his utility function	(4)

$$U = \begin{vmatrix} 0 & u_1 & u_2 & \cdots, & u_n \\ u_1 & u_{11} & u_{12} & \cdots, & u_{1n} \\ u_2 & u_{21} & u_{22} & \cdots, & u_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ u_n & u_{n1} & u_{n2} & \cdots, & u_{nn} \end{vmatrix}$$ is the determinant of the first and second partial derivatives of u . (5)

U_{rs}	is the cofactor of u_{rs} in U, and	(6)
μ	is the Lagrange multiplier used in maximizing (4) subject to (3).	(7)

it raises his "real income," and its effect along this channel is similar to that of an increase in income. [This is the Income Effect, given by the first term on the right-hand side of the equation.] On the other hand, it changes relative prices; and therefore apart from the change in real income, there will be a tendency to substitute the commodity whose price has fallen for other commodities. [This is the Substitution Effect given by the term X_{rs}.] The total effect on demand is the sum of these two tendencies [pp. 31–32].

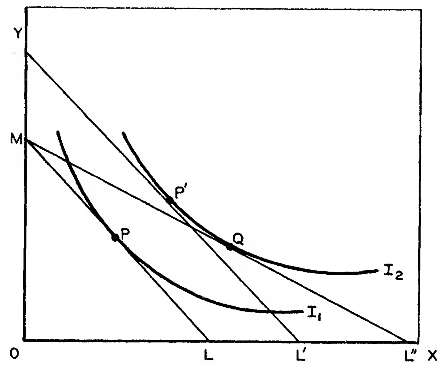


FIG. 1.—Hicks's interpretation of fundamental equation

This interpretation is illustrated by Hicks in his Figure 8 on page 31 (our Fig. 1). The curves I_1 and I_2 are two indifference curves representing two levels of utility. The initial equilibrium position is given by the point P where the budget line LM is tangent to I_1. When the price of X falls, as shown by the line $L''M$, the consumer moves to the new equilibrium position Q. This shift from P to Q is split up by Hicks into two parts; (1) the movement from P to P' and (2) the movement from P' to Q. The first represents the income effect of the change in price, or the change in demand resulting from the fact that the consumer is now better off. The second represents the substitution effect along the new indifference curve, or the change in demand resulting from the change in relative prices, apart from any change in real income.

We might equally well have considered the substitution effect first and the income effect second. In this case X_{rs} would measure the variation in demand along the *original* indifference curve to a new point S, and the income effect would measure the change in demand from S to Q associated with the change in real income from I_1 to I_2.[3] (See our Figure 2).

[3] In fact, in his investigation of the substitution effect in the theory of production, Hicks measures the substitution effect along the *original* production

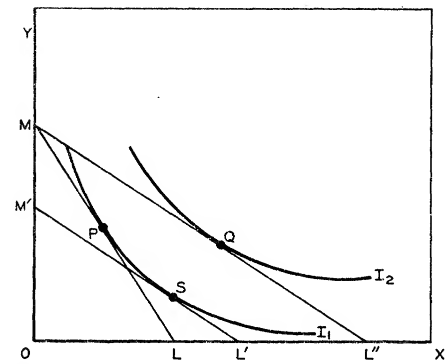


FIG. 2.—Variant of Hicks's interpretation

This conclusion has an important consequence for the theory of cost of living index numbers, since it permits us to separate the demand function into that component which is due to keeping real income unchanged and that which is due to the change in real income. By taking the first component only, we can determine the change in the cost of living corresponding to any given standard of living.

Slutsky's interpretation of the terms in equation (1) differs, however, from that of Hicks. In the words of Slutsky:

---

curve (*op. cit.*, p. 91, Fig. 21). That X_{rs} measures the substitution along the original indifference curve can easily be seen by differentiating with respect to p_r the equations $u_r = \mu p_r$ along a given indifference curve, so that $u = f(X_1, X_2, \cdots, X_n) =$ a constant. We then get

$$u_1 \frac{\partial X_1}{\partial p_r} + u_2 \frac{\partial X_2}{\partial p_r} + \cdots + u_n \frac{\partial X_n}{\partial p_r} = 0\,,$$

$$-p_1 \frac{\partial \mu}{\partial p_r} + u_{11} \frac{\partial X_1}{\partial p_r} + u_{12} \frac{\partial X_2}{\partial p_r} + \cdots + u_{1n} \frac{\partial X_n}{\partial p_r} = 0\,,$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$-p_r \frac{\partial \mu}{\partial p_r} + u_{r1} \frac{\partial X_1}{\partial p_r} + u_{r2} \frac{\partial X_2}{\partial p_r} + \cdots + u_{rn} \frac{\partial X_n}{\partial p_r} = \mu\,,$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$-p_n \frac{\partial \mu}{\partial p_r} + u_{n1} \frac{\partial X_1}{\partial p_r} + u_{n2} \frac{\partial X_2}{\partial p_r} + \cdots + u_{nn} \frac{\partial X_n}{\partial p_r} = 0\,.$$

Solving these equations, we get

$$\frac{\partial X_s}{\partial p_r} \; (u \text{ constant}) = \mu \frac{U_{rs}}{U} = X_{rs}\,.$$

If price increases by dp_r, the value $X_r dp_r$ can be said to be an *apparent loss* [in real income], since, in order to make possible the purchase of the same quantities of all the goods that had formerly been bought, the income should have to increase by $dM = X_r dp_r$. But the individual, though having the possibility of preserving unchanged the preceding [bundle of goods] will no longer consider it preferable to any other, and there will take place some kind of *residual variation* of demand:

$$\left.\begin{aligned} dX_s &= \frac{\partial X_s}{\partial p_r} dp_r + \frac{\partial X_s}{\partial M} dM , \\ &= \left(- X_r \frac{\partial X_s}{\partial M} + \mu \frac{U_{rs}}{U} \right) dp_r + \frac{\partial X_s}{\partial M} X_r dp_r , \\ &= \mu \frac{U_{rs}}{U} dp_r = X_{rs} dp_r . \end{aligned}\right\} \tag{2}$$

The increment dp_r of price, accompanied by an increment of income equal to the apparent loss can be said to be the *compensated variation* of price. In such a case X_{rs} can be regarded as [the] residual variation of demand for each unit of the compensated increment of price, and can be called the *residual variability* of X_r.[4]

The first term of equation (1), therefore, represents the additional change in demand which arises from the fact, that money income is not adjusted to compensate for the variation in price but remains constant.

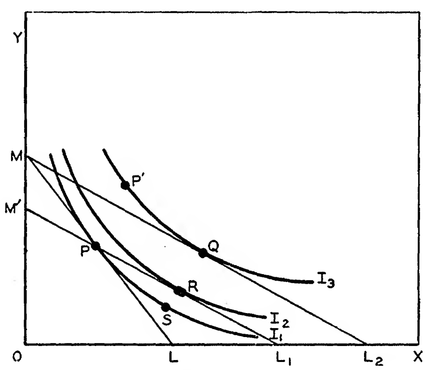


FIG 3.—Slutsky's interpretation

The meaning of these terms as Slutsky interprets them was shown very clearly by Henry Schultz in his Figure 4 of *The Theory and Measurement of Demand,* page 44. In order to compare Slutsky's interpretation with that of Hicks, we give a similar chart here (Fig. 3),

[4] Slutsky, *op. cit.,* p. 14. The quotation is taken from the translation of Slutsky made by Mr. Carl at Henry Schultz's request (unpublished manuscript, p. 21). The notation has been slightly simplified.

except that we let p_x rather than p_y vary. The initial equilibrium position is again given by the point P and the final equilibrium position by Q. But Slutsky breaks up the change into the two parts from P to R and from R to Q, rather than from P to S and from S to Q. Graphically, the point R is given by the point of tangency of an indifference curve I_2 and a line L_1M, which is parallel to L_2M and passes through P, the original equilibrium position.

The difference between the Hicks and Slutsky interpretations is readily apparent. Whereas Hicks interprets X_{rs} as measuring the substitution effect that would result if money income were adjusted so as to leave real income *actually* unchanged, Slutsky interprets it as measuring the substitution effect that would result if money income were adjusted so as to leave real income *apparently* unchanged. Whereas Hicks's substitution effect involves a change in demand along a *given* indifference curve, Slutsky's "substitution effect" involves a change from a lower to a higher indifference curve. Yet surprisingly enough, both of these effects are measured by one and the same term, X_{rs}.

Since every X_{rs} can be interpreted either with Hicks as representing the substitution along a given indifference curve, or with Slutsky as the residual variation arising from a change onto a higher indifference curve, we are left with the rather paradoxical result that the higher indifference curve is at the same time the initial indifference curve. In other words the point R coincides with the point S in Figure 3.

This paradox may be resolved as follows: If the point R coincides with the point S but also lies above it, the difference between the two must be an infinitesimal of a higher order.[5] Thus a compensated vari-

[5] I am indebted to Dr. A. Wald for the following proof of this statement, which he sent me after reading the manuscript:

"Let $f(x, y)$ be the utility (or indicator) function and denote by (x_0, y_0) the co-ordinates of the point P in your diagram. [Fig. 3]. Furthermore, let p_1 be the unit price of X, p_2 the unit price of y in the original situation (budget line LM), and $p_1 + \Delta p_1$, p_2 the prices in the new situation (budget line L_2M). Denote the co-ordinates of S by $(x_0 + \Delta x, y_0 + \Delta y)$ and the co-ordinates of R by $(x_0 + \overline{\Delta x}$, $y_0 + \overline{\Delta y})$. Then we have

$$\frac{f_x(x_0 + \Delta x, y_0 + \Delta y)}{p_1 + \Delta p_1} = \frac{f_y(x_0 + \Delta x, y_0 + \Delta y)}{p_2}, \tag{1}$$

$$f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0) \tag{2}$$

$$\frac{f_x(x_0 + \overline{\Delta x}, y_0 + \overline{\Delta y})}{p_1 + \Delta p_1} = \frac{f_y(x_0 + \overline{\Delta x}, y_0 + \overline{\Delta y})}{p_2}, \tag{3}$$

ation in price which leaves the individual's real income *apparently* unchanged also leaves it *actually* unchanged except for infinitesimals of a higher order.

It is interesting to note that in the appendix Hicks also gives the Slutsky interpretation of X_{rs}.

> Since $X_r = dM/dp_r$, when M is not taken as given, but all X's and all other p's are taken as given, it follows from the equation that the substitution term represents the effect on the demand for X_s of a change in the price of X_r combined with such a change in income as would enable the consumer, if he chose, to buy the same quantities of all goods as before, in spite of the change in p_r [p. 309].

$$(p_1 + \Delta p_1)\overline{\Delta x} + p_2\overline{\Delta y} = 0 . \tag{4}$$

Equation (4) expresses the condition that the budget line $M'L_1$ goes through P. Because of (3), (4) can be written as follows:

$$f_x(x_0 + \overline{\Delta x}, y_0 + \overline{\Delta y})\overline{\Delta x} + f_y(x_0 + \overline{\Delta x}, y_0 + \overline{\Delta y})\overline{\Delta y} = 0 . \tag{4'}$$

"Using Taylor's expansion and neglecting higher powers of Δx and Δy, we get from (2)

$$f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y = 0 . \tag{2'}$$

"In the same way we get from (4') (neglecting higher powers of $\overline{\Delta x}$ and $\overline{\Delta y}$)

$$f_x(x_0, y_0)\overline{\Delta x} + f_y(x_0, y_0)\overline{\Delta y} = 0 . \tag{4''}$$

The equations (1) and (2') in Δx and Δy are identical to the equations (3) and (4″) in $\overline{\Delta x}$ and $\overline{\Delta y}$. Since Δx, Δy, $\overline{\Delta x}$, $\overline{\Delta y}$ are of the order Δp_1, the solution of (1) and (2') differs from the solution of (1) and (2) only by a quantity of the order Δp_1^2, and the solution of (3) and (4″) differs from that of (3) and (4) also by a quantity of the order Δp_1^2. Hence $(\overline{\Delta x} - \Delta x)$ and $(\overline{\Delta y} - \Delta y)$ will be of the order Δp_1^2."

CONSTANCY OF THE MARGINAL UTILITY OF INCOME

PAUL A. SAMUELSON
Massachusetts Institute of Technology

IN THE theoretical and empirical literature on the theory of consumer's demand behavior one repeatedly encounters the hypothesis that the marginal utility of income or money can be assumed to be constant, or at least "sensibly" so.[1] As a result, many of the conclusions derived are of restricted validity; indeed, outright contradictions often emerge from this seemingly innocent assumption. It is proposed here to examine briefly, but exhaustively, the rigorous implications of the constancy of the marginal utility of income.

According to a well-known condition of equilibrium in the theory of consumer's behavior, a consumer with a given monetary income, I, confronted with respective prices, $(p_1, \cdots, p_n)$, of n goods, $(x_1, \cdots, x_n)$, will purchase each good up to the point where its marginal (degree of) utility is proportional to its price. Thus, in equilibrium

$$\phi_i(x_1, \cdots, x_n) = m_\phi p_i, \tag{1}$$

where ϕ_i represents the partial derivative with respect to the ith good of the cardinal index of utility, $\phi(x_1, \cdots, x_n)$, and m_ϕ is a factor of proportionality. This relation may be rewritten in the form

$$\begin{aligned} m_\phi = \frac{\phi_1(x_1, \cdots, x_n)}{p_1} = \frac{\phi_2(x_1, \cdots, x_n)}{p_2} = \cdots \\ = \frac{\phi_n(x_1, \cdots, x_n)}{p_n}. \end{aligned} \tag{2}$$

This proposition is often translated to mean that in equilibrium the utility of the last dollar spent in every use must be equal; the value of this magnitude is m_ϕ and will be known hereafter as the

[1] Alfred Marshall, *Principles of Economics* (8th ed.; London, 1930); Vilfredo Pareto, *Manuel d'économie politique* (Paris, 1909), Appendix; Henry Schultz, *The Theory and Measurement of Demand* (Chicago: University of Chicago Press, 1938), chap. xviii; Milton Friedman, "Professor Pigou's Method for Measuring Elasticities of Demand from Budgetary Data," *Quarterly Journal of Economics*, L(1935), 151–63; A. C. Pigou, Milton Friedman, and N. Georgescu-Roegen, "Marginal Utility of Money and Elasticities of Demand," *Quarterly Journal of Economics*, L (1935), 532–39; E. B. Wilson, "Pareto versus Marshall," *Quarterly Journal of Economics*, LIII (1939), 645–49.

marginal utility of income for the cardinal utility index, ϕ. The subscript ϕ is appended in order to emphasize its dependence upon the particular choice of utility index. It will further depend upon the equilibrium position attained by the consumer; i.e., on the set of prices and income with which he is confronted. Given such prices and income, the consumer will select a given amount of each and every good. His demand functions can be written in the form

$$x_i = h^i(p_1, \cdots, p_n, I) , \qquad (i=1, \cdots, n) \tag{3}$$

where

$$\sum_{i=1}^{n} p_i x_i = I . \tag{4}$$

One feature of these demand functions should be noted. A simultaneous doubling of all prices and income will leave the individual with the same choice as between commodities and hence will leave unaffected the amount demanded of each good. More generally,

$$h^i(\lambda p_1, \cdots, \lambda p_n, \lambda I) = h^i(p_1, \cdots, p_n, I) , \quad (i=1, \cdots, n) \tag{5}$$

where λ is any positive constant. Mathematically, the demand functions are homogeneous of order zero.

Thus, the marginal utility of income, as well as depending upon the particular cardinal index of utility selected, is also a function of all prices and income:

$$\begin{aligned} m_\phi &= \frac{\phi_i[h^1(p_1, \cdots, p_n, I), h^2(p_1, \cdots, p_n, I), \cdots]}{p_i} , \\ &= m_\phi(p_1, \cdots, p_n, I) . \end{aligned} \tag{6}$$

DEPENDENCE UPON THE CHOICE OF UTILITY INDEX

It is well known that the demand functions or schedules of the consumer remain unaltered under any change of utility index. One may write the conditions of equilibrium so that they are independent of the choice of utility index, or even by the use of indifference varieties so that no index is employed. The last approach introduces the minor inconvenience of a notational asymmetry, but this is a small price to pay for the clarity of thought which the method often imparts.

The marginal utility of income does not possess this invariance. Let us consider a transformation of the utility index of the form

$$F = F[\phi(x_1, \cdots, x_n)] = F(x_1, \cdots, x_n) , \tag{7}$$

where $F'(\phi) > 0$, and

$$
\begin{aligned}
F_i &= F'\phi_i \\
F_{ij} &= F'\phi_{ij} + F''\phi_i\phi_j \,.
\end{aligned}
\tag{8}
$$

Obviously,

$$
\begin{aligned}
m_F &= \frac{F_i}{p_i} = \frac{F'(\phi)\,\phi_i}{p_i} = F'(\phi)\,m_\phi \\
&= m_F(p_1\,, \cdots, p_n\,, I)\ .
\end{aligned}
\tag{9}
$$

In general, the change in this magnitude with respect to any parameter, say α, is also dependent upon the choice of utility index since

$$
\frac{dm_F}{d\alpha} = \frac{d}{d\alpha}\,(F'm_\phi) = F'\frac{dm_\phi}{d\alpha} + F''m_\phi\frac{d\phi}{d\alpha}\,. \tag{10}
$$

In our transformation of any utility index, no restriction can be placed upon $F''(\phi)$. Therefore, provided that $d\phi/d\alpha$ does not vanish, by an appropriate selection of utility index the rate of change in the marginal utility of income with respect to any parameter can be made to be of any arbitrary magnitude and sign. Thus, the change in the marginal utility of income with respect to a change in income, prices being constant, cannot be said to be unambiguously of any given sign. There exists an infinity of indexes for which it is of either algebraic sign at any position of equilibrium; at such a point an index can be found for which the value of this derivative is zero.

Care must be taken not to misinterpret these remarks. There need not, and in general there will not, exist any utility index for which this derivative vanishes at *every* level of income and all levels of prices. Moreover, in general there will not exist an index for which the marginal utility is constant with respect to two different parameters simultaneously. The force of these remarks will appear in the course of the argument.

There exists one particular compound change in price and income which does present an interesting invariance. Scrutiny of equation (10) reveals that

$$
\frac{1}{m_F}\frac{dm_F}{d\alpha} = \frac{1}{m_\phi}\frac{dm_\phi}{d\alpha} = \frac{d\log m}{d\alpha}\,; \tag{11}
$$

provided that $d\phi/d\alpha = 0$. That is, *the logarithmic derivative of the marginal utility of income with respect to any parameter which leaves the level of utility unchanged is invariant under any transformation of utility index.* A compound change in any single price, p_i, accompanied by a change in I sufficiently great to leave the consumer at the

same level of utility must necessarily fulfil this requirement. We should expect, therefore, to find the above coefficient invariant under such a "compensated price change."[2] This compensated price change can be written

$$\frac{1}{m_\phi}\frac{dm_\phi}{d\alpha}=\frac{1}{m_\phi}\left(\frac{\partial m_\phi}{\partial p_i}+x_i\frac{\partial m_\phi}{\partial I}\right). \tag{12}$$

Actually, by simple differentiation of equation (1), one can verify the following equivalence,[3]

$$\frac{1}{m_\phi}\left(\frac{\partial m_\phi}{\partial p_i}+x_i\frac{\partial m_\phi}{\partial I}\right)=-\frac{\partial x_i}{\partial I}. \qquad (i=1,2,\cdots,n) \tag{13}$$

The right-hand term, being a property of the demand functions, is independent of the choice of utility index so that our invariance is verified.

This also confirms the previous assertion that we cannot arbitrarily fix the sign of the change in the marginal utility of income with respect to both income and a given price, since these two must add up in a certain way (as given in eq. [13]) to a given number. Similarly, the change in the marginal utility of income with respect to each of two prices cannot be arbitrarily preassigned, since we can deduce from equation (13) the relation

$$\frac{1}{m_\phi}\left(\frac{\partial m_\phi}{\partial p_i}x_j-\frac{\partial m_\phi}{\partial p_j}x_i\right)=-\left(\frac{\partial x_i}{\partial I}x_j-\frac{\partial x_j}{\partial I}x_i\right). \tag{14}$$

The right-hand side of this equation cannot be altered by a transformation of the utility index.

In what follows reference will be made repeatedly to equation (13). It must be emphasized that it holds regardless of any assumptions of independence of utility, etc.

ALTERNATIVE INTERPRETATIONS OF CONSTANCY

The statement that the marginal utility of income is assumed to be constant is ambiguous.[4] With respect to what is it assumed to be constant? With respect to price changes? Income change? Or with respect to all of these?

[2] See Schultz, *op. cit.*, pp. 41–46.
[3] *Ibid.*, p. 40, eq. (3.15).
[4] One must be on guard against a superficial error. The fact that the marginal utility of income changes but slightly with respect to *small* changes in the variables under consideration does not imply that the *rate* of change of marginal utility with respect to these variables is small. The former proposition is a consequence simply of continuity and differentiability and holds for all functions possessing these properties.

Actually, it can be shown by simple mathematical argument that the latter cannot possibly hold. Recall that

$$m_\phi = m_\phi(p_1, p_2, \cdots, p_n, I) = \frac{\phi_i(x_1, x_2, \cdots, x_n)}{p_i}. \tag{15}$$

Consider a simultaneous doubling of all prices and income. This will leave the quantities of all goods unchanged and hence will not affect the numerator of the right-hand member. However, the denominator will be doubled. The total effect is to halve the marginal utility of income. Mathematically, m_ϕ is a homogeneous function of order minus one in the prices and income. As a consequence,

$$m_\phi(p_1, p_2, \cdots, p_n, I) = \lambda m_\phi(\lambda p_1, \lambda p_2, \cdots, \lambda p_n, \lambda I), \tag{16}$$

where λ is any positive constant. Applying Euler's theorem on homogeneous functions, we have the following identity:

$$-m_\phi \equiv \frac{\partial m_\phi}{\partial p_1} p_1 + \frac{\partial m_\phi}{\partial p_2} p_2 + \cdots + \frac{\partial m_\phi}{\partial p_n} p_n + \frac{\partial m_\phi}{\partial I} I. \tag{17}$$

Of course, m_ϕ, $p_1, \cdots, p_n$, I, all are positive quantities. Therefore, we cannot have simultaneously

$$\frac{\partial m_\phi}{\partial p_i} \equiv 0 \qquad (i = 1, 2, \cdots, n) \tag{18}$$

$$\frac{\partial m_\phi}{\partial I} \equiv 0,$$

else the right-hand sign of equation (17) would not add up to a negative quantity. This becomes even more clear if we divide (17) by m_ϕ to get the following expression involving dimensionless elasticity coefficients:

$$-1 \equiv \left(\frac{\partial m_\phi}{\partial p_1}\frac{p_1}{m_\phi}\right) + \cdots + \left(\frac{\partial m_\phi}{\partial p_n}\frac{p_n}{m_\phi}\right) + \left(\frac{\partial m_\phi}{\partial I}\frac{I}{m_\phi}\right) \tag{19}$$

The right-hand terms cannot all vanish and still add up to a -1.

How then shall we interpret constancy of the marginal utility of income? Knowingly or unknowingly, economists have formulated two distinct and alternative hypotheses under this heading. The first one involves the assumption that there exists an index of utility for which the marginal utility of income becomes independent of a change in any price, or

$$\frac{\partial m_\phi}{\partial p_i} \equiv 0. \qquad (i = 1, 2, \cdots, n) \tag{20}$$

As we have seen, this cannot possibly imply that the marginal utility of income is also constant with respect to income.

In introducing the second hypothesis attention is drawn to the fact that throughout I have repeatedly employed the term "marginal utility of *income*" rather than the "marginal utility of *money*." The latter is perhaps the term most commonly met in the literature. Not the least of its disadvantages is the fact that it suggests to the literally minded the second interpretation we are about to give.

As a preliminary it must be remembered that all demand functions and conditions of equilibrium depend upon *relative* prices and income. Very often writers arbitrarily equate the price of one good to unity, employing it as *numéraire,* or unit of reckoning, and referring to it as "money." It was only natural for such writers to conclude that constancy of the marginal utility of "money" meant constancy of the marginal utility of the good selected as *numéraire.*

Although Marshall does employ the term "marginal utility of money," there is no evidence that he had in mind this last interpretation.[5] In the first place, he rarely, if ever, employed the concept of a *numéraire.* Furthermore, he repeatedly states that the marginal utility of money decreases with income, which I shall later show to rule out the second hypothesis. He also insists that the marginal utility of money is to be associated with a flow of income rather than a stock of a commodity. This insistence is not conclusive proof but may be significant.

FIRST HYPOTHESIS

I turn then to the first or Marshallian hypothesis that the marginal utility of income is independent of price changes as expressed in equation (20). Obviously,

$$\frac{\partial m_\phi}{\partial p_i}\frac{p_i}{m_\phi} \equiv 0 \; . \qquad (i = 1, 2, \cdots, n) \tag{21}$$

Therefore, substituting into equation (19), we have

$$\frac{\partial m_\phi}{\partial I}\frac{I}{m_\phi} \equiv -1 \; . \tag{22}$$

[5] The only passages in the *Principles* which do suggest this interpretation are to be found in the Mathematical Appendix, Note XII. This argument, however, is borrowed from Edgeworth and other writers in order to establish the conclusion that market price is determinate. Against this may be cited pp. 95, 134–35, Mathematical Appendix, Note II, etc. In interpreting Marshall, it must be remembered that he regarded the constancy of the marginal utility of income as only "approximate." I shall follow out the rigorous implications of assuming this exactly; the closer is Marshall's approximation, the more nearly will the results coincide.

This says that the income elasticity of the marginal utility of income must be identically equal to minus unity. Therefore, we have

$$m_\phi = \frac{a}{I}\,, \tag{23}$$

where a is a constant. Mathematically, (23) is the only form that a homogeneous function of order -1 in a single variable can take. Equation (13) now becomes

$$\frac{\partial x_i}{\partial I}\frac{I}{x_i} \equiv -\frac{\partial m_\phi}{\partial I}\frac{I}{m_\phi} \equiv 1\,. \qquad (i = 1, 2, \cdots, n) \tag{24}$$

In words, *the seemingly innocent assumption that there exists a utility index for which the marginal utility of income is constant with respect to price changes results in the empirical restriction of unitary income elasticities of demand, or that the consumption of each and every good is exactly proportional to income.* In this case the demand functions take the special form

$$x_i = I\psi^i(p_1, \cdots, p_n) \equiv I\lambda\psi^i(\lambda p_1, \cdots, \lambda p_n)\,. \tag{25}$$

$$(i = 1, 2, \cdots, n)$$

There is a vast amount of budgetary statistical data relating to income variations in consumption. As far as I know, every investigation contradicts flatly this basic assumption. Moreover, these relations do not hold approximately even in the neighborhood of a single point.[6]

Let us go still further and investigate the strict implications of imposing an additional Marshallian assumption, namely, that there exists a utility index which is the sum of the independent utilities of each good:

$$\phi(x_1, \cdots, x_n) = f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n)\,. \tag{26}$$

Then

$$\frac{f'_i(x_i)}{p_i} = \frac{a}{I}\,. \tag{27}$$

[6] It is only in the empirically unimportant case of expenditure proportionality that the correct integrability conditions

$$\frac{\partial x_i}{\partial p_j} + x_j\frac{\partial x_i}{\partial I} \equiv \frac{\partial x_j}{\partial p_i} + x_i\frac{\partial x_j}{\partial I} \tag{1}$$

can be replaced by the special relations

$$\frac{\partial x_i}{\partial p_j} \equiv \frac{\partial x_j}{\partial p_i}\,. \tag{2}$$

Since the right-hand side of (27) does not contain $p_j (\neq p_i)$, the quantity demanded of the ith good must depend only upon its own price and income. Differentiating partially the budget equation with respect to p_j, we find

$$\frac{\partial I}{\partial p_j} = 0 = \sum_{i=1}^{n} p_i \frac{\partial x_i}{\partial p_j} + x_j \, . \tag{28}$$

Under the present independence of demand we derive

$$\frac{\partial x_j}{\partial p_j} \frac{p_j}{x_j} \equiv -1 \, . \qquad (j = 1, 2, \cdots, n) \tag{29}$$

In words, *the combined assumptions of constancy of the marginal utility of income and independence of utility imply that the elasticity of demand be always unity.* This together with the assumption that all demand functions be homogeneous of order zero requires that the demand functions take the special form

$$x_i = k_i \frac{I}{p_i} \, , \qquad (i = 1, 2, \cdots, n) \tag{30}$$

where the k's are constants equal to the fraction of income spent on each commodity.

Equation (27) can be rewritten

$$f'_i \left(k_i \frac{I}{p_i}\right) = a \frac{p_i}{I} \, . \tag{31}$$

Therefore,

$$f'_i (v) = \frac{a k_i}{v} \, , \tag{32}$$

or

$$f'_i (x_i) = \frac{a k_i}{x_i} \, . \qquad (i = 1, 2, \cdots, n) \tag{33}$$

Hence,

$$\phi = b + a (k_1 \log x_1 + k_2 \log x_2 + \cdots + k_n \log x_n) \, . \tag{34}$$

This result could also have been derived with the aid of the generalized form of Burk's theorem that expenditure proportionality plus independence of utilities restricts severely the indifference surfaces and utility indexes.[7]

As an illustration of the lack of invariance of the marginal utility

[7] A. Burk, "Real Income, Expenditure Proportionality, and Frisch's 'New Methods of Measuring Marginal Utility,'" *Review of Economic Studies*, IV, No. 1 (1936), 33–52.

of income the reader is invited to compute this magnitude for the transformed, equivalent utility index

$$F = B + Ax_1^{k_1} x_2^{k_2} \cdots x_n^{k_n} . \tag{35}$$

It need hardly be said that no empirical observations justify the imposition of such a definite form upon the utility indexes of indifference surfaces and demand functions.

SUFFICIENCY CONDITIONS FOR CONSTANCY IN THE FIRST SENSE

The introduction of the assumption of independence of utilities was in the nature of a digression designed to show the damage that may result from the addition of seemingly harmless and plausible hypotheses. I return now to the hypothesis of constancy of the marginal utility of income. We have seen that it implies (income) *expenditure proportionality.* Is this all? Are there perhaps still further hidden restrictions which more thorough investigation might reveal?

It is desirable to be able to answer definitely such questions. Fortunately, it can be proved that no further restrictions are implied by this hypothesis. This can be done once and for all by proving that the necessary conditions of equation (25) are also sufficient to insure the existence of a utility index in terms of which the marginal utility of income is constant with respect to price changes. The proof consists of a specification of the steps by means of which such an index can be constructed.

If expenditure proportionality holds, the indifference surface can be written in the following special form:

$$\frac{F'(\phi)\phi_i}{F'(\phi)\phi_j} = {}^jR^i(x_1 , x_2 , \cdots , x_n) = {}^jR^i(\lambda x_1 , \lambda x_2 , \cdots , \lambda x_n) ; \tag{36}$$

i.e., the slope functions are homogeneous of order zero. This being the case, a utility index can be found which is homogeneous of the first order in the quantities of goods,

$$\Phi = \Phi(x_1 , x_2 , \cdots , x_n) = \frac{1}{\lambda} \Phi(\lambda x_1 , \lambda x_2 , \cdots , \lambda x_n) . \tag{37}$$

In fact, the indifference surfaces can be "numbered" according to their respective distances from the origin along an arbitrary "spoke" going through the origin, and we will have such an index. In terms of this index the conditions of equilibrium can be written

$$\Phi_i = m_\Phi p_i . \quad (i = 1, 2, \cdots , n) \tag{38}$$

Multiplying each equation by x_i, respectively, and adding, we have

$$\sum_{i=1}^{n} \Phi_i x_i = m_\Phi \sum_{i=1}^{n} p_i x_i \; . \tag{39}$$

Because of the budget equation (4) and the fact that Φ obeys Euler's theorem for homogeneous functions of the first order, this can be written

$$m_\Phi = \frac{\Phi}{I} \; . \tag{40}$$

Consider the utility index, ϕ, defined as

$$\phi = b + a \log \Phi \; . \tag{41}$$

From equation (9) we find

$$m_\phi = \frac{a}{\Phi} m_\Phi = \frac{a}{I} \; , \tag{42}$$

which is identically equation (23). Hence, expenditure proportionality implies that a utility index can always be found for which the marginal utility of income is independent of prices. This completes the sufficiency proof and guarantees that there are no hidden additional implications of the assumption of constancy of the marginal utility of income in the Marshallian sense.

CONSTANCY OF THE MARGINAL UTILITY OF "MONEY"

I now turn to a brief analysis of the empirical implications of the second possible interpretation of constancy of the marginal utility of income.[8] According to this point of view, the marginal utility of some one good, which can be designated as *numéraire,* is constant with respect to changes in all *other* prices and income. Designating the *numéraire* as the first good, we have

$$\begin{aligned} \frac{\partial m_\phi}{\partial p_i} &\equiv 0 \; ; \qquad (i = 2, 3, \cdots, n) \\ \frac{\partial m_\phi}{\partial I} &\equiv 0 \; . \end{aligned} \tag{43}$$

Because m_ϕ is homogeneous of order -1 in all prices and income, we have

[8] See Wilson, *op. cit.;* J. R. Hicks, *Value and Capital* (Oxford: Clarendon Press, 1939), pp. 38–41.

$$m_\phi \equiv \frac{a}{p_1}\,. \tag{44}$$

Substituting equations (43) into equation (13), we have

$$\begin{aligned} &\frac{\partial x_i}{\partial I} \equiv 0\,, \qquad (i=2, 3, \cdots, n) \\ &\frac{\partial x_1}{\partial I} \equiv -\frac{\partial m_\phi}{\partial p_1}\frac{1}{m_\phi} \equiv \frac{1}{p_1}\,. \end{aligned} \tag{45}$$

The demand functions must take the special form[9]

$$\begin{aligned} &x_i = G^i(p_1, \cdots, p_n) = G^i(\lambda p_1, \cdots, \lambda p_n)\,, \qquad (i=2, 3, \cdots, n) \\ &x_1 = \frac{I}{p_1} - \sum_{i=2}^{n} \frac{p_i}{p_1} G^i(p_1, \cdots, p_n)\,. \end{aligned} \tag{46}$$

This means that any increase in income is spent completely on one commodity. It need hardly be said that all empirical budgetary studies show this hypothesis to be absurd.

As with the first hypothesis, I shall now prove the sufficiency of the necessary conditions of equations (46). Because of the homogeneity condition on the G's, the first $(n-1)$ equations can be solved to give the prices, $(p_2/p_1,\ p_3/p_1, \cdots, p_n/p_1)$ in terms of the goods, $(x_2, x_3, \cdots, x_n)$. These price ratios are independent of x_1, so the indifference curves take the special form

$$\frac{p_i}{p_1} = R^i(x_2, x_3, \cdots, x_n)\,. \qquad (i=2, 3, \cdots, n) \tag{47}$$

The linear differential expression

$$a\,dx_1 + \sum_{i=2}^{n} aR^i dx_i\,, \tag{48}$$

is an exact differential because the existence of at least one utility index implies

$$R_j^{\,i} - R_i^{\,j} \equiv R^j R_1^{\,i} - R^i R_1^{\,j} \equiv 0\,. \tag{49}$$

Hence, a utility index exists of the form

$$\phi = b + a x_1 + a \sum_{i=2}^{n} \int_{c_i}^{x_i} R^i(x_2, \cdots, x_n)\,dx_i\,. \tag{50}$$

[9] Pigou appears to be in error in his belief that constancy of the marginal utility of income with respect to income implies infinite elasticity of demand. Equations (46) reveal no such implication.

For this index,

$$m_\phi = \frac{a}{p_1} \ ,$$

which is identically equal to equation (44). This proves the sufficiency of the conditions of equation (46).

As with the previous case, it is of interest to consider the simultaneous effects of the assumption of independence of utilities and constancy of the marginal utility of income. In this case, a utility index exists of the following form

$$\phi = b + ax_1 + f_2(x_2) + f_3(x_3) + \cdots + f_n(x_n) \ . \tag{51}$$

Hence,

$$m_\phi = \frac{f'_i(x_i)}{p_i} = \frac{a}{p_1} \ . \tag{52}$$

But

$$x_i = G^i\left(1 \ , \frac{p_2}{p_1} \ , \frac{p_3}{p_1} \ ; \cdots \ , \frac{p_n}{p_1}\right) \tag{53}$$

so that

$$f'_i\left[\ G^i \ \left(1 \ , \frac{p_2}{p_1} \ , \frac{p_3}{p_1} \ , \cdots \ , \frac{p_n}{p_1}\right)\right] = a\frac{p_i}{p_1} \ . \tag{54}$$

Since the right-hand side does not involve (p_j/p_1) $(\neq p_i/p_1)$, neither can the left, requiring our demand curves to take the special form

$$\begin{aligned} x_i &= G^i\left(\frac{p_i}{p_1}\right), \qquad (i = 2, 3, \cdots, n) \\ x_1 &= \frac{I}{p_1} - \sum_{i=2}^{n} \frac{p_i}{p_1} G^i\left(\frac{p_i}{p_1}\right). \end{aligned} \tag{55}$$

This does *not* imply unitary elasticity of demand, as some earlier writers have thought. In fact, unitary elasticity would lead to a contradiction. Of course, the restrictions imposed by the combined assumptions are even more incompatible with the facts of economic life.

CONSUMER'S SURPLUS AND ALLIED CONCEPTS

We have seen the empirical implications of the various hypotheses with respect to the strict constancy of the marginal utility of income and noted their incompatibility with statistical observations. It is desirable before concluding to indicate the effects of this analysis upon various constructions which depend for their validity upon the constancy of the marginal utility of income.

The first of these is Marshallian consumer's surplus. Before examining it in detail, let us consider the uses to which it is put. Among other things it is proposed as a measure of the gain (loss) of utility that results from a decrease (increase) in price of a single good. An attempt also has been made to apply it to the analysis of the burden involved in commodity taxation. It has been used to determine the maximum amount of revenue that a perfectly discriminating monopolist might exact from the consumer for a given amount of the good in question.

Since only an ordinal preference field is assumed in the theory of consumer's behavior, there is really little importance to be attached to any numerical measure of the gains from a price change. In particular, one cannot fruitfully compare the gain derived from a movement between two given price situations with the gain between two other price situations.[10] Moreover, all valid theorems relating to the burden of taxation can be stated independently of any numerical measure of utility change. We should not be greatly perturbed, therefore, if the concept of consumer's surplus should be found to be inadmissible. Its only advantage seems to lie in its easy two-dimensional representation.

Consider an initial price and income situation, $(p_1^a, \cdots, p_n^a, I^a)$, and the corresponding amount of goods purchased, $(x_1^a, \cdots, x_n^a)$. For any selected utility index, ϕ, there will also be a given amount of utility, $\phi(X^a)$. Suppose that a change is made in but one price, p_i, and income is left unchanged. There will be new amounts of every commodity, $(x_1^b, \cdots, x_n^b)$, and of utility, $\phi(X^b)$, corresponding to the new prices and income, $(p_1^b, \cdots, p_n^b, I^b)$, or $(p_1^a, p_2^a, \cdots, p_i^b, \cdots, p_n^a, I^a)$.

We are interested in the following magnitudes:

1. The gain (loss) in utility resulting from the price change, or $\phi(X^b) - \phi(X^a)$.

2. The area between the demand curve of the ith good and the p_i axis within the range of the price movement, or

$$- \int_{p_i^a}^{p_i^b} x_i dp_i \, .$$

3. The amount by which the expenditure on the ith good in the new situation is exceeded by the maximum amount of money which

[10] One can, however, compare the gains derived from a change in the basic price situation with an alternative price change from the *same* basic situation, since this resolves itself into an *ordinal* comparison of the alternative new situations. The initial situation "cancels out" so to speak.

the consumer would be willing to pay for x_i^b in preference to trading at the old set of prices. (This may be negative if we are dealing with a price increase rather than a decrease.) Call this E_{ab}.

4. The change in income which will make trading at the new set of prices as attractive as trading at the old set of prices with the initial income. Call this ΔI_{ab}.

5. The change in income which will make trading at the old set of prices as attractive as trading at the new set of prices with the initial income. Call this $\Delta I'_{ab}$.[11]

According to the Marshallian doctrine of consumer's surplus, all five of these magnitudes are equal except for dimensional constants. We are explicitly warned, however, that this doctrine holds unqualifiedly only when the marginal utility of income is constant, and only if utilities are independent. I shall now examine the value of each of these magnitudes in four cases: (a) in the general unrestricted case of stable demand; (b) under the first interpretation of constancy of the marginal utility of income; (c) under the second hypothesis when the ith good is not the *numéraire*; and (d) under the second hypothesis when the ith good itself has constant marginal utility of income. Only the most sketchy proofs will be indicated.

In the general case we have the following relations:

$$\begin{aligned}\phi(X^b) - \phi(X^a) &= \int_{p_i^a}^{p_i^b} \left(\frac{d\phi}{dp_i}\right) dp_i = \int_{p_i^a}^{p_i^b} \sum_{j=1}^{n} \left(\frac{\partial\phi}{\partial x_j}\frac{\partial x_j}{\partial p_i}\right) dp_i \\ &= \int_{p_i^a}^{p_i^b} m_\phi \sum_{j=1}^{n}\left(p_j \frac{\partial x_j}{\partial p_i}\right) dp_i = -\int_{p_i^a}^{p_i^b} m_\phi\, x_i\, dp_i .\end{aligned} \tag{56}$$

$$\Delta I_{ab} = \max\left(\sum_{j=1}^{n} p_j^b x_j^b - \sum_{j=1}^{n} p_j^b x_j\right), \quad \text{where } \phi(X) = \phi(X^a), \tag{57}$$

$$\geqq \sum_{j=1}^{n} p_j^b (x_j^b - x_j^a), \tag{58}$$

$$\geqq \sum_{j=1}^{n} (p_j^a - p_j^b) x_j^a . \tag{59}$$

If only the ith price changes, this becomes

$$\Delta I_{ab} \geqq (p_i^a - p_i^b) x_i^a . \tag{60}$$

Similarly,

$$\Delta I'_{ab} \leqq \sum_{j=1}^{n} p_j^a (x_j^b - x_j^a) \tag{61}$$

[11] Note that $\Delta I_{ab} = -\Delta I'_{ba}$; $\Delta I'_{ab}$, but not ΔI_{ab} nor E_{ab}, can exceed I.

$$\leqq \sum_{j=1}^{n} (p_j^a - p_j^b) x_j^b \tag{62}$$

and

$$\Delta I'_{ab} \leqq (p_i^a - p_i^b) x_i^b . \tag{63}$$

It is impossible in the general case[12] to determine the relative magnitude of ΔI_{ab} and $\Delta I'_{ab}$. Hence,

$$\Delta I_{ab} \gtreqless \Delta I'_{ab} . \tag{64}$$

It can be shown that

$$E_{ab} = -\int_{p_i^a}^{p_i^b} x_i dp_i + \int_{p_i^a}^{p_i^b} (\bar{p}_i - p_i) \frac{\partial x_i}{\partial p_i} dp_i , \tag{65}$$

where $\bar{p}_i$ is the price which would have to prevail for the consumer *freely* to select the batch of goods which he actually does consume when presented with an "all-or-none" offer by the perfectly discriminating monopolist. The first term on the right-hand side of equation (65) is the area under the demand curve. The second "correction" term may be of either sign.[13] It also follows from the definition of ΔI_{ab} that

$$\Delta I_{ab} \geqq E_{ab} . \tag{66}$$

In case (b) we find

$$E_{ab} < \Delta I_{ab} < \Delta I'_{ab} , \tag{67}$$

and

$$\frac{\phi(X^b) - \phi(X^a)}{m_\phi} = -\int_{p_i^a}^{p_i^b} x_i dp_i > E_{ab} .^{14} \tag{68}$$

The following relations must be satisfied in case (c):

$$\frac{\phi(X^b) - \phi(X^a)}{m_\phi} = -\int_{p_i^a}^{p_i^b} x_i dp_i = E_{ab} = \Delta I_{ab} = \Delta I'_{ab} . \quad (i \neq 1) \tag{69}$$

[12] If we rule out the inferior good phenomenon so that demand is "normal,"

$$\Delta I_{ab} < \Delta I'_{ab} . \tag{1}$$

Actually,

$$\Delta I_{ab} - \Delta I'_{ab} = \int_{p_i^a}^{p_i^b} \{x_i(p_1^a, p_2^a, \cdots, p_i, \cdots p_n^a, \phi^b)$$

$$- x_i(p_1^a, p_2^a, \cdots, p_i, \cdots, p_n^a, \phi^a)\} dp_i .$$

[13] In the "normal" two-dimensional case it will be of negative sign; i.e., a perfectly discriminating monopolist will be able to exact less than the area under the demand curve from the consumer.

[14] The last of these inequalities will certainly hold in the two good case. I have not developed a satisfactory proof that it holds for the n-dimensional case.

Although this is not the Marshallian interpretation, consumer's surplus seems to be most justified in this case. However, the above equalities cannot hold simultaneously for every good.[15]

For case (*d*) we have the same relations as case (*b*); i.e., equations (67) and (68) must hold with the possible exception of the inequality referred to in a previous footnote.[16]

[15] Case (*c*) is *sufficient* to insure the equalities of eq. (69). Some of them may hold in other cases.

[16] It is of some interest to calculate these magnitudes in the *purest Marshallian case* when utilities are independent and the marginal utility of income is independent of price changes. Here we have for consumer's surplus

$$\begin{aligned}\frac{\phi(X^b)-\phi(X^a)}{m_\phi} &= -\int_{p_i^a}^{p_i^b} x_i\,dp_i\,,\\ &= k_i I \log\frac{p_i^a}{p_i^b} > 0\,,\quad p_i^a > p_i^b\,.\end{aligned}\tag{1}$$

If we compute consumer's surplus from the origin, i.e., from the price at which the consumption of x_i is equal to zero, we find that consumer's surplus is infinite! That is

$$\lim_{p_i^a\to\infty}\log\frac{p_i^a}{p_i^b}=\infty\,.\tag{2}$$

Some writers regard consumer's surplus as being infinite but suggest that it may be made finite if measured from some minimum or subsistence level of the commodity in question. In the purest Marshallian case since the demand curve approaches the price axis asymptotically in such a way that the integral or area under the curve is divergent, there exists no such unique minimum level.

Also

$$E_{ab}=I\left\{1-\left(\frac{p_i^b}{p_i^a}\right)^{k_i/(1-k_i)}\right\}\leqq I\,,\tag{3}$$

$$\Delta I_{ab}=I\left\{1-\left(\frac{p_i^b}{p_i^a}\right)^{k_i}\right\}\leqq I\,,\tag{4}$$

$$\Delta I'_{ab}=I\left\{\left(\frac{p_i^a}{p_i^b}\right)^{k_i}-1\right\}>\Delta I_{ab}\,.\tag{5}$$

In the limit as p_i^a goes to infinity we have

$$E_{ab}=I=\Delta I_{ab}\,,\tag{6}$$

$$\Delta I'_{ab}=\infty\,.\tag{7}$$

When the Marshallian assumptions are met perfectly, the area under the demand curve is not equal to the amount that a perfectly discriminating monopolist could exact from the consumer by an all or none offer—even though the marginal utility of income is constant!

HISTORICAL SUMMARY

In concluding I should like to touch briefly on the history of the discussion of the marginal utility of income. As cited earlier, Marshall assumed constancy in the first sense as a basis for his doctrine of consumer's surplus. He also employed constancy in the second sense to show that market price is determinate. It is amusing in this connection to notice that he did not feel it necessary to point out that the accidental determinateness of final price in this case is accompanied by complete indeterminateness in the quantity of one of the goods.

Pareto disliked the constancy of the marginal utility of income, although it is not possible to infer unambiguously from his *Manuel* in which sense he interpreted this. He clearly considers it to be constant with respect to $(n-1)$ prices; since he does not differentiate with respect to the price of the *numéraire* good, either interpretation might hold. Throughout he deals only with the case of independent utilities. An incomplete derivation of the theorem of unitary price elasticity of demand for each good under these conditions suggests that it would be necessary to adopt the first interpretation.

Assuming marginal utility of income to be constant, utilities to be independent, and the proportion spent on each good to be small relative to the income elasticities of demand, Pigou suggested that the ratio of the price elasticities of demand of two goods is equal to the ratio of their respective income elasticities of demand.[17] The latter can be computed from budgetary studies. By the use of a theorem established by Friedman[18] it can be shown that this same conclusion follows from the second and third of these assumptions and is independent of the assumption of constancy of the marginal utility of income. It still seems to depend upon independence of utilities, the empirical implication of which I deal with in a forthcoming paper.

Aside from its use in connection with consumer's surplus, the main concern of writers has been to derive the negative slope of the demand curve by means of the assumption of constancy. For it is only when marginal utility of income is constant and utilities are independent that the slope of the demand curve can be derived from the behavior of the curve of marginal utility. In this case, the alleged law of diminishing marginal utility implies negatively sloping demand curves. We have seen here that the rigorous assumption of this relation for every good implies something much more definite; namely, unitary price and income elasticity of demand and vanishing cross-elasticities of demand.

[17] A. C. Pigou, *Economics of Welfare* (London, 1920), Appen. II.
[18] *Op. cit.*

A CONTRIBUTION TO THE NONSTATIC THEORY OF PRODUCTION*

GERHARD TINTNER [1]
Iowa State College

IN WHAT follows I wish to consider some aspects of the theory of production under conditions which are nonstatic (in the sense of Hicks).[2] These problems involve time and hence anticipations. We can assume these anticipations as single valued or as merely probable. If the probability distribution of the expectations is considered to be known with certainty (probability one), we will talk of *subjective risk.* If the probability distribution of the anticipations is not itself known definitely but only anticipated again with a certain likelihood, we will talk about *subjective uncertainty.* We add the word "subjective" since these notions are related to, but by no means identical with, Knight's ideas about risk and uncertainty.[3] We do not here consider any "objective" risk, but only the subjective estimate and opinion of the individual.[4] Hence our concept of probability is here the same as Keynes,[5] and different from, for example, the frequency definition of von Mises.[6] A possible connection between these two approaches if the individual bases his subjective probability estimate upon the objective frequencies experienced in the past will be considered in the last section.

* This article will be filed as Journal Paper No. J-895 of the Iowa Agricultural Experiment Station, Ames, Iowa. Project No. 557.

[1] The author is greatly indebted to A. G. Hart (Iowa State College), O. Lange (University of Chicago), D. Rockefeller (New York City), T. W. Schultz (Iowa State College), and Mary Wise (University of Chicago) for advice and criticism.

[2] J. R. Hicks, *Value and Capital* (Oxford, 1939), pp. 115 ff.

[3] F. H. Knight, *Risk, Uncertainty and Profit* (Boston, 1921 [reprint London, 1933]), pp. 197 ff.

[4] E. Lindahl, *Studies in the Theory of Money and Capital* (London, 1939), pp. 348 ff. See also P. N. Rosenstein-Rodan, "The Role of Time in Economic Theory," *Economica* (N.S.), I (1934), 77 ff.; O. Morgenstern, "Vollkommene Voraussicht und wirtschaftliches Gleichgewicht," *Zeitschrift für Nationalökonomie,* VI (1935), 337 ff.

[5] J. M. Keynes, *Treatise on Probability* (London, 1921).

[6] R. von Mises, *Probability, Statistics and Truth* (London, 1939). See also E. Nagel, *Principles of the Theory of Probability* ("International Encyclopedia of Unified Science," Vol. I, No. 6 [Chicago, 1939]).

I. STATICS

Following mainly Hicks[7] and Allen [8] we will quickly review the static, timeless case, in order to develop a notation which will later be adapted to the nonstatic case. Let $x_1, x_2, \cdots, x_m$, be the (positive) quantities of commodities produced and the negative quantities of factors used in the production of a firm. Let the transformation function be

$$f(x_1, x_2, \cdots, x_m) = 0\,. \tag{1}$$

We consider here for simplicity's sake only the case of *one* transformation function. The complication of more than one such relationship (limitational factors)[9] will be introduced later.

The profit which the firm is trying to maximize is

$$V = \sum_{v=1}^{m} x_v p_v\,, \tag{2}$$

where p_v is the price of the commodity or the factor of production number v. We will assume that these prices are determined on the market independently of the actions of the firm in question. The generalization to the case of monopoly and monopolistic competition is, however, obvious.[10]

The problem of maximizing the sum (2) under the condition (1) is equivalent to maximizing a new function $V^* = V - \mu f$, where μ is a Lagrange multiplier. The firm can maximize the profit under the given conditions by buying the appropriate amounts of the factors of production and producing appropriate amounts of the products. In other words, it can choose the x_v. Differentiating, we get for the necessary conditions of a maximum:

$$\mu \frac{\partial f}{\partial x_v} = \mu f_v = p_v. \qquad (v = 1, 2, \cdots, m) \tag{3}$$

It follows that, for example, the marginal productivities are proportional to the prices. These m equations (3) together with equation

[7] *Op. cit.*, pp. 78 ff., 319 ff.

[8] R. G. D. Allen, *Mathematical Analysis for Economists* (London, 1938), pp. 502 ff. See also E. Schneider, *Theorie der Produktion* (Vienna, 1934), J. L. Mosak, "Interrelations of Production, Price and Derived Demand," *Journal of Political Economy*, XLVI (1938), 761 ff., and S. Carlson, *A Study in the Pure Theory of Production* (London, 1939).

[9] N. Georgescu-Roegen, "Fixed Coefficients of Production and Marginal Productivity Theory," *Review of Economic Studies*, III (1935), 40 ff.

[10] J. R. Hicks, "Annual Survey of Economic Theory: The Theory of Monopoly," *Econometrica*, III (1935), 4 ff.; J. Robinson, *Economics of Imperfect Competition* (London, 1933), pp. 235 ff.

(1) are sufficient to determine unambiguously the $m+1$ values $x_1, x_2, \cdots, x_m, \mu$, if some other conditions are fulfilled which are analogous to those pointed out by Wald[11] and by von Neumann.[12]

The second-order conditions assure us that we have a true maximum rather than a minimum or a minimax.[13] They can be stated in terms of the determinant:

$$F = \begin{vmatrix} 0, f_1, f_2, \cdots, f_n \\ f_1, f_{11}, f_{12}, \cdots, f_{1n} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ f_n, f_{1n}, f_{2n}, \cdots, f_{nn} \end{vmatrix}, \tag{4}$$

where

$$\frac{\partial^2 f}{\partial x_u \partial x_v} = f_{uv}.$$

It must be negative together with all its principal minors in order to give a true maximum of profit.

We can solve the set of equations (3) which together with (1) determine the quantities bought or sold as functions of the prices:

$$x_v = x_v(p_1, p_2, \cdots, p_m). \qquad (v = 1, 2, \cdots, m) \tag{5}$$

Differentiating equation (1) and the system (3), we get a set of linear equations in terms of the demand or supply differentials dx_u:

$$\sum_{v=1}^{m} f_v dx_v = 0 \tag{6}$$

$$f_u d\mu + \mu \sum_{v=1}^{m} f_{uv} dx_v = dp_u. \qquad (u = 1, 2, \cdots, m)$$

This gives for the demand or supply differential the following very general expression:

$$dx_u = \frac{\sum_{v=1}^{m} F_{uv}\, dp_v}{F \mu}, \qquad (u = 1, 2, \cdots, m) \tag{7}$$

where F_{uv} is the cofactor of the element f_{uv} in the determinant F (formula [4]).

[11] A. Wald in *Ergebnisse eines mathematischen Kolloquiums*, ed. Karl Menger (1935, 1936); "Über einige Gleichungssysteme der mathematischen Ökonomie," *Zeitschrift für Nationalökonomie*, VII (1936), 637 ff., G. Stigler, *Production and Distribution Theories* (New York, 1941), 243 ff.

[12] J. von Neumann in *Ergebnisse eines mathematischen Kolloquiums*, ed. Menger (1938).

[13] Allen, *op. cit.*, pp. 495 ff.; Hicks, *Value and Capital*, pp. 302 ff., 319 ff.

For the price elasticities of demand for the factors of production and elasticities of supply of the products we have, by specialization of (7),

$$\frac{Ex_v}{Ep_u}=\frac{F_{uv}\,p_u}{\mu F x_v}=K_v\,\sigma_{uv}\,, \tag{8}$$

where $K_v=p_vx_v/V$ is the relative expenditure on or revenue from v and $\sigma_{uv} = (F_{uv}\sum_{w=1}^{m} x_wf_w)/(x_ux_vF)$ is the "elasticity of substitution" between u and v. This is an elasticity of substitution in the narrower sense if u and v are both commodities produced or are both factors of production. Otherwise it is an elasticity of transformation.

Let $d_v = p_vx_v$ be the expenditure upon v (if v is a factor of production) or the revenue from v (if v is a commodity produced). We have, then, for the price elasticities of expenditure or revenue if $u \neq v$:

$$\frac{Ed_v}{Ep_u}=\frac{Ex_v}{Ep_u}\,, \qquad (u\neq v) \tag{9}$$

since the elasticity of a constant multiple of a function is the same as the elasticity of the function itself.

If, however, $u = v$, we have

$$\frac{Ed_u}{Ep_u}=\frac{Ex_u}{Ep_u}+1. \tag{10}$$

We can, finally, consider not just one firm but the whole economy. Let ${}_jx_v$ be the demand or supply of the commodity v by the firm (or private individual) number $j\,(j = 1, 2, \cdots, M)$.

We have for the total demand or supply for v

$$X_v=\sum_{j=1}^{M} {}_jx_v=X_v\,, \tag{11}$$

which is again a function of the prices. If we introduce the share of the firm number j in the total demand or supply of v, ${}_jL_v = {}_jx_v/X_v$, we have for the price elasticities of the total demand or supply of the factor or product v:

$$\frac{EX_v}{Ep_u}=\sum_{j=1}^{M}\frac{E_jx_v}{Ep_u}\,{}_jL_v \tag{12}$$

if $u \neq v$, and

$$\frac{EX_u}{Ep_u}=\sum_{j=1}^{M}\frac{E_jx_v}{Ep_u}\,{}_jL_v+1 \tag{13}$$

if $u = v$.

II. DYNAMICS: ANTICIPATIONS SINGLE VALUED[14]

We consider first the discontinuous case and assume that the individual or the firm plans over a time interval $1, 2, \cdots, n$. We assume a fixed economic horizon (Tinbergen).[15] Denote the amount of the commodity u which is anticipated to be produced at the point in time t or the negative amount of the factor u anticipated to be bought at the point in time t as x_{ut}. (Continuous input and output will be considered later.) Let i_t be the anticipated rate of interest at the point in time t and $r_t = 1 + i_t$ the anticipated accumulation rate. Denote further by $R_t = r_1, r_2, \cdots, r_t$ the total accumulation rate for the whole interval $1, 2, \cdots, t$. Let p_{ut} be the anticipated price of the commodity or factor u at the point in time t and $q_{ut} = p_{ut}/R_t$ the anticipated discounted price. All anticipations are single valued.

The individual or firm will try to maximize the total anticipated discounted profit W:

$$W = \sum_{v=1}^{m} \sum_{s=1}^{n} x_{vs} q_{vs} \tag{14}$$

under the condition of a transformation function which now involves all the anticipated factors and products over the whole interval $1, 2, \cdots, n$:

$$g(x_{11}, x_{12}, \cdots, x_{mn}) = 0 . \tag{15}$$

(The generalization for the case of more than one transformation function is obvious.) This is again the same problem as to maximize a new function $W^* = W - \mu g$. We get for solutions:

$$\mu \frac{\partial g}{\partial x_{ut}} = \mu g_{ut} = q_{ut}, \quad (u = 1, 2, \cdots, m, t = 1, 2, \cdots, n) \tag{16}$$

and these mn equations together with the conditional equation (15) are sufficient to determine the planning of production, that is, to find the $mn + 1$ quantities x_{ut}, μ. This only holds true as before if some other conditions are fulfilled. We could also state the second-order conditions in a way similar to the static case. The determinant

[14] See especially Hicks, *Value and Capital*, pp. 115 ff.; G. Tintner, "The Maximization of Utility over Time," *Econometrica*, VI (1938), 154 ff.; "The Theoretical Derivation of Dynamic Demand Curves," *ibid*, pp. 375 ff.; "Elasticities of Expenditure in the Dynamic Theory of Demand," *ibid.*, VII (1939), 266 ff.; "Some Remarks on the Dynamic Theory of Production," in Cowles Commission for Research in Economics, *Report of the Fifth Annual Research Conference* (Chicago, 1939), pp. 61 ff. See also G. L. S. Shackle, *Expectations, Investment and Income* (New York, 1938).

[15] J. Tinbergen, "The Notion of Horizon and Expectancy in Dynamic Economics," *Econometrica*, I (1933), 247 ff.

$$G = \begin{vmatrix} 0\,, & g_{11}\,, & g_{12}\,, & \cdots, g_{mn} \\ g_{11}\,, & g_{11,11}\,, & g_{11,12}\,, & \cdots, g_{11,mn} \\ g_{mn}\,, & g_{11,mn}\,, & g_{12,mn}\,, & \cdots, g_{mn,mn} \end{vmatrix}, \tag{17}$$

[where now $g_{ut,vs} = (\partial^2 g)/(\partial x_{ut}\, \partial x_{vs})$] must be negative and also its principal minors.

By an argument which is analogous to the one presented for the static case we get the following expression for the demand or supply differentials for factor u or commodity u anticipated to be demanded or supplied at the point in time t:[16]

$$dx_{ut} = \frac{\sum_{v=1}^{m} \sum_{s=1}^{n} G_{ut,vs}\, dq_{vs}}{\mu\, G}\,. \tag{18}$$

We have to remember that according to its definition $\partial\, q_{vs}/\partial\, p_{zw} = 0$ for $z \neq v$, $s \neq w$, and $= 1/R_s$ otherwise. Also $\partial\, q_{vs}/\partial\, r_w = 0$ for $s < w$ and $= -\ q_{vs}/R_w$ otherwise.[17] Remembering these relations, we can easily derive from the general formula (18) price and accumulation rate derivatives and elasticities.

The price elasticities for the anticipated demand for factors of production or of the anticipated supply of products are very similar to the static case:

$$\frac{Ex_{vs}}{Ep_{ut}} = \frac{G_{vs,ut}\, p_u}{\mu\, Gx_{vs}} = K_{vs}\, \sigma_{ut,vs}\,, \tag{19}$$

where now $K_{vs} = q_{vs}\, x_{vs}/W$ is the relative expenditure on, or revenue from, the factor or commodity v at the point in time s. But

$$\sigma_{ut,vs} = (G_{ut,vs} \sum_{z=1}^{m} \sum_{w=1}^{n} x_{zw}\, g_{zw}) / (x_{ut}\, x_{vs}\, G)$$

is the elasticity of substitution or transformation of factor or commodity u at the point in time t for factor or commodity v at the point in time s.

The accumulation rate elasticity of demand or supply of factor or commodity v at the point in time s with respect to the anticipated accumulation rate at the point in time t, r_t is:

$$\frac{Ex_{vs}}{Er_t} = -\sum_{w=t}^{n} \sum_{z=1}^{m} K_{zw}\, \sigma_{vs,zw} = \sum_{w=1}^{t-1} \sum_{z=1}^{m} K_{zw}\, \sigma_{vs,zw}. \tag{20}$$

[16] Tintner, "Some Remarks, ... ," *op. cit.*, p. 62.
[17] Tintner, "The Theoretical Derivation . . . . ," *op. cit.*, p. 377.

It hence appears as a weighted sum of elasticities of substitution. Elasticities of expenditure or revenue can be derived in much the same way as with the theory of demand.[18]

The results for the market demand or supply are similar to the static case assuming that everybody acts according to anticipations. We define: ${}_jx_{vs}$ is the demand or supply of the firm j anticipated for commodity v at the point in time s. Adding up the results of the business plants of the M firms, we have:

$$X_{vs} = \sum_{j=1}^{M} {}_jx_{vs} = X_{vs}(p_{11}, \cdots, p_{mn}, r_1, \cdots, r_n), \tag{21}$$

and we can derive price and interest elasticities of market demand and supply in a similar fashion as for the static case.

We abandon now the discontinuous case and turn to continuous production. The anticipated inputs and outputs $x_v(s)$, $v = 1, 2, \cdots, m$, vary now continuously over the interval of planning $0 \le s \le n$. The transformation function g becomes a functional $\bar{g}$

$$\bar{g} = \bar{g} \begin{bmatrix} \overset{n}{\underset{0}{x_1(s)}}, \overset{n}{\underset{0}{x_2(s)}}, \cdots \overset{n}{\underset{0}{x_m(s)}} \end{bmatrix}. \tag{22}$$

The discounted anticipated net profit to be maximized is now

$$\bar{W} = \sum_{v=1}^{m} \int_0^n x_v(s) q_v(s) ds. \tag{23}$$

The solution yields a system of m functional equations:

$$\lambda \bar{g}'_v[x_1, x_2, \cdots, x_m; s] = q_v(s), \qquad (v = 1, 2, \cdots, m, 0 \le s \le n) \tag{24}$$

where $\bar{g}'$ is the functional derivative with respect to v at the point in time s and λ is a Lagrange multiplier.

Functional demand and supply derivatives and elasticities can be derived from a system of integral equations similarly to the ones occurring in the theory of consumption.[19]

III. SUBJECTIVE RISK

We understand under subjective risk the following situation: A firm is again at the point in time 0 and plans for the interval 1, 2, $\cdots$, n (discontinuous case). Its anticipations of some noncontrollable

[18] Tintner, "Elasticities of Expenditure . . . . ," *op. cit.*
[19] Tintner, "The Theoretical Derivation . . . . ," *op. cit.*, p. 380.

factors[20] are now not assumed to be single valued, but there is a *probability distribution* specifying the probabilities which the firm assigns to various values of all these factors at different points in time during the planning interval $1, 2, \cdots, n$.[21] This probability distribution itself is supposed to be known *definitely*.

More specifically, let us use again the notation of the beginning of the previous section. Let us make further the simplifying assumption that only the anticipations with respect to the future prices and interest rates are not single valued and that the transformation function itself is anticipated with certainty.[21a]

The net discounted profit of the firm is

$$W = \sum_{v=1}^{m} \sum_{s=1}^{n} x_{vs}\, q_{vs}\,, \tag{25}$$

but $q_{vs} = p_{vs}/R_s$ is not now known definitely since the future prices (p_{vs}) and accumulation rates (r_s) are not themselves anticipated with certainty. Let us assume that the probability distribution which expresses the *subjective risk anticipations* of the firm is given by

$$\begin{aligned} Pdp_{11}, dp_{12}, \cdots, dp_{mn}, dr_1, dr_2, \cdots, dr_n & \\ = P(p_{11}, p_{12}, \cdots, p_{mn}, r_1, r_2, \cdots, r_n; k_1, k_2, \cdots, k_K) & \\ dp_{11}, dp_{12}, \cdots, dp_{mn}, dr_1, dr_2, \cdots, dr_n; & \end{aligned} \tag{26}$$

that is, the probability that simultaneously p_{11} will lie in the interval between p_{11} and $p_{11} + dp_{11}$, p_{12} in the interval between p_{12} and $p_{12} + dp_{12}, \cdots$, p_{mn} in the interval between p_{mn} and $p_{mn} + dp_{mn}$, r_1 in the interval between r_1 and $r_1 + dr_1$, r_2 in the interval between r_2 and $r_2 + dr_2, \cdots$, r_n in the interval between r_n and $r_n + dr_n$. The quantities $k_1, k_2, \cdots, k_K$ are K parameters which characterize the probability distribution P. They are, for instance, the means and variances of the expected prices and accumulation rates, as well as the correlation coefficients between prices and accumulation rates, and

[20] Tintner, "A Note on Economic Aspects of the Theory of Errors in Time Series," *Quarterly Journal of Economics*, LIII (1938), 141 ff.

[21] Hicks, *Value and Capital*, pp. 125 ff. See also "The Theory of Uncertainty and Profit," *Economica*, XI (1931), 170 ff.; "A Suggestion of Simplifying the Theory of Money," *ibid.* (N. S.), I (1935), 1 ff.; A. C. Pigou, *The Economics of Welfare* (2nd ed.; London, 1924), Appen. I, pp. 725 ff.; A. G. Hart, "Anticipations, Business Planning and the Cycle," *Quarterly Journal of Economics*, LI (1937), 15 ff.; Lindahl, *op. cit.*, pp. 40 ff;. A. G. Hart, *Anticipations, Uncertainty, and Dynamic Planning*, ("Studies in Business Administration," Vol. XI, No. 1, [Chicago, 1940]).

[21a] For a discussion of the case of technical and technological risk and uncertainty see my forthcoming article, "The Pure Theory of Production under Technological Risk and Uncertainty," *Econometrica*, IX (1941), in print.

higher population moments or cumulants of the multivariate probability distribution.

We can easily derive from this probability distribution (26) the probability distribution of the anticipated net profit W, for example, by the method of characteristic functions.[22] This probability distribution is

$$QdW = Q(W; x_{11}, x_{12}, \cdots, x_{mn}, k_1, k_2, \cdots, k_K)\, dW, \qquad (27)$$

i.e. the probability that the discounted net profit W will lie in the interval between W and $W + dW$. This probability still depends upon the controllable quantities x_{vs}—upon the anticipated inputs and outputs, as well as upon the parameters of the probability distribution P, k_j.

We must now define a preference functional, say ϕ, which expresses the way in which the individual or the firm in question evaluates the probabilities attached to specific discounted net profits. Following the work of Menger[23] and Marschak,[24] we want a very general statement and hence write:

$$\phi = \phi\left[Q(\overset{+\infty}{\underset{-\infty}{W}})\right]. \qquad (28)$$

That is, we assume a risk preference functional which is fixed by the total shape of the probability distribution of profit, Q, over its whole range. This includes the special case in which the individual tries to maximize the mathematical expectation of profit (where $\phi = \int_{-\infty}^{+\infty} WQ[W]dW$) which has already been discussed by Jevons.[25] It is also valid if higher moments enter into the consideration, especially the coefficient of variability and the skewness of the distribution Q (Marschak).[26]

The risk preference functional ϕ has to be maximized under the condition of our old transformation function g as given in (15), since we assume that the conditions of production are anticipated with certainty. We assume again only *one* transformation function. The theory can be easily generalized for several functions.

The first-order conditions of a maximum of the risk preference

[22] See e.g., S. S. Wilks, *Statistical Inference* (Princeton, 1937), pp. 34 ff.

[23] K. Menger, "Das Unsicherheitsmoment in der Wertlehre," *Zeitschrift für Nationalökonomie*, Vol. V (1934).

[24] J. Marschak, "Money and the Theory of Assets," *Econometrica*, VI (1938), 311 ff.; H. Makower and J. Marschak, "Assets, Prices and Monetary Theory," *Economica* (N.S.), *V* (1938), 261 ff.

[25] W. S. Jevons, *The Theory of Political Economy* (4th ed.; London, 1934), pp. 72 ff.

[26] Marschak, *op. cit.*, pp. 320 ff.

functional ϕ under the condition $g = 0$, (formula [15]), are given by:

$$\int_{-\infty}^{+\infty} \phi'[Q(W); W_1] \frac{\partial Q(W_1)}{\partial x_{vs}} dW_1 = \lambda g_{vs}. \qquad (v = 1, 2, \cdots, m, s = 1, 2, \cdots, n) \tag{29}$$

Here λ is again a Lagrange multiplier and ϕ' is the functional derivative of ϕ with respect to Q at the point $W = W_1$. This together with condition (15) gives us the constant λ and also the anticipated amount of the factors bought or commodities produced at different points in time in the interval $1, 2, \cdots, n$. (x_{vs}).

The second-order conditions for the maximum of a functional with side conditions have been stated in the literature but are not of special interest for the economist.[27]

In order to investigate the influence of changes in the probability function P upon the determination of the expected inputs and outputs x_{vs}, we have to consider its parameters. It depends upon K parameters $k_1, k_2, \cdots, k_K$ (formula [26]). The probability distribution of net profits Q will also involve these parameters (formula [27]). The inputs and outputs x_{vs} will depend upon these constants.

For a general expression, we find again the total differentials of g (formula [15]) and of the system (formula [29]):

$$\sum_{v=1}^{m} \sum_{s=1}^{n} g_{vs} dx_{vs} = 0$$

$$g_{ut}\, d\lambda + \lambda \sum_{v=1}^{m} \sum_{s=1}^{n} g_{vs,ut}\, dx_{vs} = d \int_{-\infty}^{+\infty} \phi'[Q(W); W_1] \frac{\partial Q(W_1)}{\partial x_{ut}} dW_1 \quad . \qquad \begin{pmatrix} u = 1, 2, \cdots, m \\ v = 1, 2, \cdots, n \end{pmatrix} \tag{30}$$

Solving this system of equations, we get for the demand or supply differential:

$$dx_{ut} = \frac{\sum_{v=1}^{m} \sum_{s=1}^{n} G_{ut,vs}\, d \int_{-\infty}^{+\infty} \phi'[Q(W); W_1] \frac{\partial Q(W_1)}{\partial x_{vs}} dW_1}{\lambda G}. \tag{31}$$

Specializing, we find, for instance, for the derivative with respect to k_j:

[27] See, e.g., H. H. Goldstine, "The Minima of Functionals with Associated Side Conditions," *Duke Mathematical Journal*, III (1937), 418 ff.

$$\frac{\partial x_{ut}}{\partial k_j} = \frac{\sum_{v=1}^{m}\sum_{s=1}^{n} G_{ut,vs}}{\lambda G} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\{\phi'[Q(W);W_1,W_2]\frac{\partial Q(W_1)}{\partial x_{vs}}\frac{\partial Q(W_2)}{\partial k_j} \qquad (32)$$
$$+\phi'[Q(W);W_1]\frac{\partial^2 Q(W_1)}{\partial x_{vs}\,\partial k_j}\}\,dW_1\,dW_2.$$

This is the derivative of x_{ut}, of the anticipated production of commodity u or demand for factor u at the point in time t with respect to the parameter k_j. For instance, the latter may be the mean of a price or an accumulation rate, or its variance or a correlation coefficient between two prices, etc., in the distribution P. We can investigate in this way the influence of the characteristics of the joint probability distribution P as given in formula (26) upon the planned economic action of the firm in the interval $1, 2, \cdots, n$.

The continuous case presents no new problems. We have simply continuous outputs and inputs in the interval $(0,n)$ instead of discontinuity. As a result of the maximization of a risk preference functional the planned inputs and outputs $x_v(s)$ $(0 \leq s \leq n)$ appear as functions of the parameters of the joint distribution function of the anticipated prices and accumulation rates in the interval $(0,n)$.

IV. SUBJECTIVE UNCERTAINTY [28]

We understand under subjective uncertainty the following: The firm is again in the point in time 0 and plans for the period $1, 2, \cdots, n$. We consider first the discontinuous case. The expectations of the noncontrollable factors—the prices and accumulation rates anticipated to prevail in this interval $1, 2, \cdots, n$—are again not anticipated with certainty, but there is a joint probability distribution, depending upon the K parameters k_j, which we may think of as of the moments of this distribution P as given in formula (26).

But this probability distribution $P = P(p_{11}, p_{12}, \cdots, p_{mn}, r_1, r_2, \cdots, r_n; k_1, k_2, \cdots, k_K)$ is now not assumed to be known definitely. This is to say, the individual in question realizes that his information is limited and assumes that the values of the parameters k_j in P are not absolutely known with certainty (i.e., probability one) but only with a likelihood depending upon past experience. We follow Marschak[29] in the choice of the term "likelihood" in analogy of certain ideas of

[28] Marschak, *op. cit.*, pp. 323 ff.
[29] *Ibid.*

R. A. Fisher.[30] Let us assume that there is a *likelihood function* for the individual, representing his knowledge of the parameters k_j:

$$Ldk_1, dk_2, \cdots, dk_K = L(k_1, k_2, \cdots, k_K; h_1, h_2, \cdots, h_H)\, dk_1, dk_2, \cdots, dk_K. \tag{33}$$

This is the likelihood with which the individual in question anticipates the values of the parameters of the probability function P to fall into certain intervals. It shows the likelihood that k_1 lies between k_1 and $k_1 + dk_1$, k_2 between k_2 and $k_2 + dk_2, \cdots, k_K$ between k_K and $k_K + dk_K$. The H new parameters $h_1, h_2, \cdots, h_H$ are now the characteristics of the distribution of the k_j. We can think of them as the parameters, for example, the moments, of the multivariate distribution L (formula [33]). All this expresses the individual's subjective uncertainty about the probability distribution P. We have really a "probability distribution of probability distributions."

Given the probability of the anticipated net profit W from formula 27 as $Q = Q(W; x_{11}, x_{12}, \cdots, x_{mn}; k_1, k_2, \cdots, k_K)$, that is, as depending upon the controllable factors x_{vs} (anticipated inputs and outputs) as well as upon the parameters k_j, which characterize the probability distribution P, we can now derive the likelihood of the anticipated discounted net profit W

$$MdW = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} QLdk_1, dk_2, \cdots, dk_K\, dW = M(W; h_1, h_2, \cdots, h_H)\, dW. \tag{34}$$

This is the likelihood of the discounted expected profit W to lie between W and $W + dW$, derived from the likelihood of the distribution P itself. It still depends upon the H parameters h_j which characterize this likelihood L.

We introduce a new preference functional ψ which expresses the preference for subjective uncertainty

$$\psi = \psi[M(W)]. \tag{35}$$

ψ depends now upon the likelihood of expected net profit (M). It is only fixed if this function is given over its entire range.

The functional (35) must be a maximum under the condition of formula (15), $g = 0$. The generalization for more than one transfor-

[30] *Statistical Methods for Research Workers* (3d ed.; London, 1930), pp. 243 ff. Our likelihood is however more in the nature of an a priori probability (see below).

mation function is obvious. This yields the following set of conditions

$$\int_{-\infty}^{+\infty} \psi' \, [M;W_1] \, \frac{\partial M(W_1)}{\partial x_{vs}} \, dW_1 = \lambda \, g_{vs} \,, \qquad (v = 1, 2, \cdots, m, s = 1, 2, \cdots, n) \tag{36}$$

where λ is a Lagrange multiplier and ψ' is the functional derivative of ψ with respect to M at the point $W = W_1$. The expected inputs and outputs x_{vs} appear now as functions of the parameters h_j of the likelihood function L. Derivatives and elasticities with respect to these parameters can be derived in a similar fashion as before.

V. FLEXIBILITY AND ADAPTABILITY

A. G. Hart[31] and G. Stigler[32] have pointed out the importance of a consideration of questions of flexibility and adaptability of production processes for planning. Marschak[33] has stressed the plasticity of assets. These problems arise only in the case which we called subjective uncertainty. The individual in question does not know definitely the probability distribution of anticipated discounted net profits but has only a likelihood estimate of the possible probability distributions themselves. Hence he may believe that he can improve his knowledge as time goes on and will try to keep as flexible or adaptable as possible. He will have to commit himself for the near future but try to keep reasonably flexible and adaptable for the more distant future. He believes that there is an advantage for him to have later the possibility to change or to modify his production plans upon the basis of new information (Marschak).

It is not necessary to make here the distinction between flexibility and adaptability, where the latter has certain absolutely fixed resources which are fully exploited. We simply assume again that a firm is at the point in time 0 and plans for the discontinuous points $1, 2, \cdots, n$. Anticipations of prices and accumulation rates are single valued. Using the symbols introduced in Section II, we have the total discounted anticipated net profit

$$W = \sum_{v=1}^{m} \sum_{s=1}^{n} x_{vs} \, q_{vs} \,.$$

The individual tries to maximize it. But we have not only one condi-

[31] *Op. cit.*, pp. 21 ff.; "Imputation and Demand for Productive Resources in Disequilibrium," in *Explorations in Economics* (New York, 1936), pp. 264 ff.

[32] "Production and Distribution in the Short Run," *Journal of Political Economy*, XLVII (1939), 305 ff.

[33] *Op. cit.*, pp. 321 ff.

tion (transformation function), $g = 0$, as before, but assume a number of such conditions. There exist, say, p relationships between the anticipated inputs and outputs at various points in time x_{vs}, say

$$g_1 = 0, \quad g_2 = 0, \cdots, g_p = 0.$$

We hence define a new function

$$W^* = W - \sum_{i=1}^{p} \mu_i g_i, \tag{37}$$

which he has to maximize. We get for the first-order conditions of a maximum:

$$\sum_{i=1}^{p} \mu_i g_{i,ut} = q_{ut}, \qquad (u = 1, 2, \cdots, m, t = 1, 2, \cdots, n) \tag{38}$$

where now $g_{i,ut} = \partial g_i / \partial x_{ut}$. This is the complete solution if we assume single-valued anticipations for the prices and accumulation rates.

If, however, we abandon this assumption, we have again a probability distribution P as in formula (26) for the anticipated prices p_{vs} and the anticipated accumulation rates r_s. Assume this probability distribution to depend upon K parameters k_j. This is the case of subjective risk. But in the case of subjective uncertainty there is also a likelihood function L (formula [33]) which indicates the likelihood of specific values of the parameters k_j and hence special forms of the probability distribution P. We assume now that the individual in question believes that he will be able to improve his forecasts in the future. Hence he wants to retain some flexibility for his future economic actions.

This may be expressed in the following way. Assume the existence of a second-order likelihood function N for the parameters of L

$$N = N(h_1, h_2, \cdots, h_H; a_1, a_2, \cdots, a_A), \tag{39}$$

which takes the possibility of later changes in the plans into consideration. It gives the likelihood of various likelihood functions L by giving the probability of their parameters h_j. The a_j are parameters, again the moments and product moments of the distribution N. We can derive a new likelihood function for the expected discounted net profit W

$$\begin{aligned} BdW = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} MNdh_1\, dh_2 \cdots dh_H\, dW \\ = B(W; a_1, a_2, \cdots, a_A)\, dW. \end{aligned} \tag{40}$$

We have to define a new preference functional for the second-order likelihood of profit B:

$$\chi = \chi[B(W)]. \tag{41}$$

The maximum conditions are now

$$\int_{-\infty}^{+\infty} \chi'[B\,;\,W_1]\,\frac{\partial B(W_1)}{\partial x_{vs}}\,dW_1 = \sum_{i=1}^{n} \mu_i\, g_{i,vs} \qquad (v = 1, 2, \cdots, m\,,\ s = 1, 2, \cdots, n) \tag{42}$$

where again the μ are Lagrange multipliers and χ' is the functional derivative of χ with respect to B at the point $W = W_1$. This gives the sufficient conditions for the situation where adaptability and flexibility are important.

VI. MODELS OF ANTICIPATIONS

Single-valued anticipations.—A number of writers have investigated the way in which anticipations of future economic events could possibly be influenced by past experience.[34] I want to mention in this connection especially some indication given by Keynes[35] and the work of some other writers, for example, Hicks,[36] as well as the earlier work of Evans[37] and of Roos.[38] Assuming single-valued anticipations, we want to derive a number of models which are of the simplest type, mathematically speaking linear functionals.

The simplest assumption, of course, is the following: the individual projects the present value of some economic factors into the future. This is to say, for instance, that the present price is anticipated to prevail also in the future. A somewhat more complicated model would consist in projecting the simple average of some factor over a given past interval into the future. A certain refinement could be introduced by using not a simple but a weighted average (Roos, Evans).[39] It is pretty evident that the weights would become the smaller the farther the events to which they pertain lie back

[34] G. Tintner, "A Note on Economic Aspects of the Theory of Errors in Time Series," *op. cit.*, esp. p. 145.

[35] J. M. Keynes, *The General Theory of Employment, Interest and Money* (New York, 1936), pp. 46 ff., 147 ff.

[36] "The Meeting of the Econometric Society in Leyden, September–October 1933," *Econometrica*, II (1934), 194 f.; *Value and Capital*, pp. 125 ff.

[37] G. C. Evans, *Mathematical Introduction to Economics* (New York, 1930), pp. 143 ff.

[38] C. F. Roos, *Dynamic Economics* (Bloomington, Ind., 1934), pp. 14 ff.

[39] Evans, *op. cit.;* Roos, *op. cit.;* H. T. Davis, *The Theory of Econometrics* (Bloomington, Ind., 1941), chap. xvii, pp. 377 ff.

of functionals which are known as linear integral equations.
in the past. Mathematically speaking, these models belong to the type

Another type of assumption which we can make is that not simply the prevailing value of the factor but its past *trend* is projected into the future and forms the basis of anticipations. Mathematically speaking, this introduces the first and possibly higher time derivatives. Not only the slope but also the curvature and possibly other geometrical characteristics of the trend may in this way be projected into the future and form a basis of anticipations. This type of relationship would yield us the type of functionals which appears in the calculus of variations.

Another type of relationship which may be the basis of the formation of anticipations is the following: There is one particular point of time in the past which will influence the anticipations for the future whereas the time before and after has no or little influence. For instance, we can imagine that a war or a great crisis will have more influence on the formation of certain types of anticipations than whatever has happened apart from them. If only the value at certain past points in time influences the formation of anticipations we get, mathematically speaking, a type of functionals which is connected with the theory of difference equations (Frisch, Kalecki, Tinbergen).[40]

Subjective risk.—With subjective risk we have essentially the situation of the games of chance.[41] We assume a probability distribution of anticipations for the future, which probability distribution itself, however, is known with certainty (probability one). Different individuals will have different preferences for high or low probabilities, according to the question if they are more or less gamblers. A gambler, for instance, will prefer long odds, this is to say, a large gain even if connected with a low probability. This introduces the property of the skewness of the probability distribution. But also the mathematical expectation, this is to say, the gain which can be anticipated on the average in the long run; and the variance, this is to say, the spread of the possible values of gain and loss about the average will be important, etc. (Menger, Marschak).[42] We can assume that there is for every individual a preference function or rather preference functional which will depend on the probability distribution mentioned.

[40] R. Frisch, "Propagation and Impulse Problems in Dynamic Economics," in *Economic Essays in Honor of Gustav Cassel* (London, 1933), pp. 171 ff.; J. Tinbergen, *Statistical Testing of Business-Cycle Theories* (2 vols.; Geneva, 1939); M. Kalecki, *Essays in the Theory of Economic Fluctuations* (London, 1939).

[41] Marschak, *op. cit.*, pp. 324 f.

[42] Menger, *op. cit.*, Marschak, *op. cit.*

We can assume that the probability distribution will depend on past experience much in the same way as mentioned in the previous section. For instance, the mean of the probability distribution of the anticipated factor may be its mean over some past interval. The variance and other characteristics may also be derived similarly from past experience.

Subjective uncertainty—In the case of subjective uncertainty the situation is such that *the probability distributions themselves* are not known definitely. Hence we have to consider the likelihood of these distributions themselves.[43] This problem is very similar to Bayes's theorem, which has been very extensively discussed by writers on probability.[44]

We have to establish the likelihood of probability distributions corresponding to so called "a priori distribution" in the theory of probability.[45] Its characteristics also may be based in some way on past experience of the individual. It seems to be difficult to construct models for this case. This particular feature of uncertainty seems to be important for the function of the innovator or creative entrepreneur (Schumpeter).[46] The rentier in the sense of Pareto,[47] on the other hand, has aversion against subjective risk and subjective uncertainty. It is very difficult to say in which way the entrepreneur evaluates the uncertainty of his anticipated probability distributions, and it may even seem that those evaluations are not essentially and certainly not exclusively based upon past experience. We could, however, try to construct models, for instance, on the basis of the number of successful and unsuccessful patents and new business ventures. This could possibly serve as a guide for a model for the rational conduct of the entrepreneur when dealing with genuine subjective uncertainty.

Another aspect of the problem has to do with the time sequence of economic action and its effect upon economic planning, the permanency of the decisions, and the possibility of changing them later.[48] The entrepreneur will have to make some sacrifice of anticipated profit in order to keep his enterprise more adaptable and flexible and to be able to make new decisions and change his plans if his knowledge increases in the future and uncertainty decreases (Hart, Stigler).

[43] Marschak, *op. cit.*, pp. 323 ff.

[44] See, e.g., J. V. Uspensky, *Introduction to Mathematical Probability* (New York, 1937), pp. 60 ff.

[45] *Ibid.*

[46] J. A. Schumpeter, *The Theory of Economic Development* (Cambridge, Mass., 1934), pp. 128 ff.; *Business Cycles*, I (New York, 1939), 87 ff.

[47] V. Pareto, *The Mind and Society*, IV (New York, 1935), 1555 ff.

[48] Hart, *op. cit.;* Stigler, *op. cit.*

This will imply a new likelihood which expresses his estimate of the probability that he will have to change his plans.

We can introduce a preference functional for subjective uncertainty in the sense indicated above. Whereas a gambler has a taste for subjective risk, an entrepreneur or innovator of the Schumpeterian type will be a man who prefers subjective uncertainty. He bases his action upon a probability distribution of profits which distribution has itself small likelihood but would yield large profits. It seems doubtful, however, if this case should be analyzed at all in terms of rational economic behavior. It is more likely that prestige or power are here the moving forces.

RISK, UNCERTAINTY, AND THE UNPROFITABILITY OF COMPOUNDING PROBABILITIES[1]

ALBERT GAILORD HART
Iowa State College

THE study of problems of anticipations—especially of the demand for money and the valuation of stocks—has led to intensification of interest in "risk" and "uncertainty." Both risk and uncertainty, in the terminology which now tends to become standard, are *subjective* [2] matters—attributes of anticipations and (by extension) of plans for action. "Risk" is taken to denote the holding of anticipations which are not "single valued" but constitute a probability distribution having known parameters. "Uncertainty" is taken to denote the holding of anticipations under which the parameters of the probability distribution are themselves not single valued.[3]

It is the position of this paper that "risk" has comparatively little importance in economic analysis in view of the characteristics of the time relations in which we are interested. It will be shown that, while an uncertainty situation can be described as a risk situation if we apply the rule of compound probabilities, the use of this reduction technique tends to obscure rather than to clarify the economic issues. In consequence the writer urges that theorists concentrate their attention on uncertainty rather than on risk.

2. The notion that uncertainty exists when the parameters of a probability estimate are not single valued seems to offer two avenues for exploration. One is to work out the consequences of the assumption that probability estimates are ordinal rather than cardinal: that

[1] The writer is indebted to his colleagues, Drs. Tintner, Winsor, and Kozlik, for their pains in helping to clarify both the economics and the mathematics of the following paper; but as none of them has seen the final outcome, they cannot be held liable for it.

[2] Following Myrdal, some Swedish economists distinguish between "subjective" and "objective" risks. But they mean by the latter not the "risk" as it would be estimated with full knowledge of all elements of the situation—when, of course, a given outcome would be regarded as either certain or impossible, not merely as likely—but as it would be estimated *from the data available to the actual estimator* by an ideal estimator (see E. Lindahl, *Studies in the Theory of Money and Capital* [London, 1939], pp. 348–49).

[3] On the terminological point see J. Marschak, "Money and the Theory of Assets," *Econometrica*, October, 1938, p. 324; G. Tintner, "A Contribution to the Nonstatic Theory of Production," in this volume, (pp. 92–109).

contingency A is considered more likely than contingency B, but the estimator has no definite notion how much more likely.[4] The second is to work on the assumption that all probability estimates are cardinal. This implies that the estimator is sure that some one of a set of alternative probability distributions (having different parameters) is the one by which he should plan, but that he is not sure which one, though he has estimates of their relative likelihood. That is, if we assume the estimator to make cardinal estimates, uncertainty implies that he has *a probability distribution of probability distributions.*[5]

The assumption of ordinal probabilities, however, seems to the present writer to lead into a blind alley. If probability estimates are merely ordinal, their expectation value, dispersion, skewness, etc., lack measurability; and for lack of units we are unable to set up preference scales among them. The economist's normal procedure of breaking up a planning problem into components is thus blocked. Preferences among alternative events and estimates of the relative probability of different events under different courses of action are inextricably intertwined, and we can say only that the individual does what he prefers in view of the whole constellation of circumstances. In consequence this paper will proceed on the assumption that estimated probabilities are cardinal quantities.

3. At first glance it may appear that an uncertainty situation so conceived is simply a special case of risk. If the planner has a probability distribution of probability distributions for (say) the price of firecrackers next July, all he need do is to multiply each distribution by its probability, and summate (or integrate), and the whole system of anticipations is boiled down into a single probability distribution for the price. Since this is so, why should not theorists content themselves with showing that this sort of reduction is possible and thereafter concern themselves only with risk?

The answer is that the two sorts of probabilities involved—which we may follow Tintner in calling "probabilities" under particular possible probability distributions and "likelihoods" of those distributions occurring—play very different roles in planning. Even though an entrepreneur (say) is interested solely in the expectation value of the distribution of *possible profits*—to the neglect of the standard deviation and higher moments—he must take into account the stand-

[4] Marschak is the writer who has gone farthest in the direction of trying to free himself from the assumption that probability estimates are cardinal (see his article just cited, also the alternative version published in *Economica*, August, 1938).

[5] This is the procedure of Tintner.

ard deviation and higher moments of the likelihoods existing in his estimates of *possible prices*, etc. But unless he is interested in the higher moments of the profit distribution, the dispersion of probabilities within the component distributions will be completely indifferent if the number of components distinguished is sufficiently[6] large. The merging of the various contingent probability distributions means merging the likelihoods and the probabilities by multiplication, so that the resulting total distribution conceals some of the data relevant for business or household planning.[7]

4. This difference in the bearing of likelihoods and of probabilities under component distributions turns on two economic considerations: (*a*) the anticipation of a change in anticipations and (*b*) the possibility of deferring decisions, with or without special costs. If either one of these elements is lacking in a planner's situation, his affliction is risk rather than uncertainty; nothing is gained by keeping the likelihoods and the probabilities separate. In fact, if he does not expect his anticipations to change, this implies that there is only one component distribution, with a likelihood of unity. If plans once made are to be completely inflexible, there will be no opportunity to make use of later improvements in estimates. In either case, the total distribution contains all the data he is able to use.

Both elements, however, must be assumed to exist in most planning problems. If we are able to form any estimates at all of such future magnitudes as prices, these estimates must be based upon a stock of evidence accumulated in the past. But unless the event in question is imminent, the future must be expected to bring in more relevant evidence. Possibly new evidence will change our outlook and give our estimates a radically different expectation value. More probably new evidence will confirm our impressions and leave the expectation value substantially unchanged. In this case it will almost certainly reduce the dispersion of our estimates. Such a *convergence* of anticipations is normally expected, since we ordinarily think we can estimate the nearer future more accurately than the more remote

[6] For the criterion of "sufficiency" see n. 11 below.

[7] That something is lost in the multiplication of likelihoods and probabilities is readily seen from the irreversibility of the merger. While any probability distribution of component probability distributions has but one sum, any total distribution may be the sum of any of an infinity of sets of component distributions. The parameters of the total distribution do not tell us, e.g., whether it is the sum of a large number of components each with a small dispersion, or of a small number of components each with a wide dispersion, or of a large number of components which are very much alike, each with a wide dispersion. Yet for planning purposes these are substantially different cases.

future.[8] But anticipations, from their very nature, can scarcely fail to contain an anticipation that they will change.[9]

5. Given the two conditions just described, the pattern of planning may be worked out by projecting ourselves into the situation of the latest date at which decisions affecting output of given date will be taken, and developing the plan backward toward the present.[10] Giving hypothetical values to the inputs to be determined earlier, it is possible to say what will be the input scheme offering the greatest expectation of profit in the light of each price-probability-distribution for output now considered possible; this hypothetical decision will depend merely on the expectation value of each such distribution, not at all on its dispersion or skewness. For each contingency[11] there is thus a profit expectation. Multiplying each expectation by the likelihood ascribed to the corresponding probability distribution at the date of planning gives a combined expectation.

The value of this expectation depends on the setting of the final decision, created largely by the previous inputs. If now we reopen the next-to-last input decision, it is plain that each possible determination of the next-to-last input will create a different situation for the application of the last input. Each of these situations has a corresponding profit expectation (net of outlays for the next-to-last input) ; and the next-to-last decision is in substance a choice among these situations. Plainly the highest profit expectation should be selected. By a chain

[8] Note that this does *not* necessarily imply that our estimates for very short periods beginning now are more accurate than for longer periods beginning now. On account of random elements affecting particular days, to give an example, a department-store executive can expect to estimate next week's sales with greater relative accuracy than tomorrow's. What *is* implied is that tomorrow's sales can be estimated better than those of day after tomorrow, next week's sales better than those of week after next.

Estimates for very short periods beginning now are inaccurate on account of such random elements; estimates for very long periods beginning now are inaccurate because they cannot allow properly for the drift of systematic causal elements operating through time. There must thus be some period for which accuracy is a maximum—ranging perhaps from a day to a season according to the kind of business dealing in view.

[9] It is conceivable that the list of component distributions may contain one having all parameters identical with those of the total distribution and that this component may be ascribed a high likelihood; but the case is obviously trivial.

[10] The argument which follows contains an implicit assumption that outputs of different dates are independently adopted in order to save space. The writer has published a sketch of the necessary corrections in "Anticipations, Business Planning and the Cycle," *Quarterly Journal of Economics*, February, 1937, and a fuller version in *Anticipations, Uncertainty, and Dynamic Planning* (Chicago, 1940), pp. 60 ff.

[11] The number of component probability distributions which must be factored out of the total distribution before we can neglect the dispersion of probabilities within the components is therefore whatever number is considered now (at the planning date) to be possible for the date when the last decision will be taken (cf. paragraph 3 above).

process of reasoning along these lines, decision at each stage can be formulated in the light of its effects on the setting of later decisions.[12] This process thus makes it possible to work out a whole plan, though it involves considering an enormous number of variants.

The plan thus arrived at will differ from the plan resulting when all inputs must be definitively planned at the outset in several important respects: (*a*) its expectation value of profit will be higher; (*b*) it must rest not merely on the expectation value of the selling price but on the whole distribution of likelihoods of estimates which may be held at the date of final decision; and (*c*) its dispersion of profit prospects will be different and probably smaller (at least measured by the coefficient of variation).

6. The fact that the expectation value of profit will be higher than under a rigid plan follows from the planner's freedom to make his flexible plan just like the rigid plan if he chooses. Obviously, then, the flexible plan (if chosen) cannot be inferior. But even if the initial input under the flexible plan is just what it would be under the rigid plan, the improvement of estimates by the time of the next input decision will ordinarily show an optimum second input different from the input which would have figured in the rigid plan. That is, by starting along the lines of the rigid plan and then shifting as estimates are modified, the profit can be increased. Over and above this gain, there is a possibility that an initial input different from that figuring in the rigid plan will offer a still more favorable setting for later decisions, however estimates may shift.

7. The importance of the higher moments of the estimate distributions arises from the fact that in a flexible plan the inputs beyond the initial date, as well as the output, will be uncertain—that is, will be planned hypothetically, in several variants corresponding to the different courses it is believed the evolution of estimates may follow. If estimates of the selling price are revised upward, the higher values of later inputs and of outputs will be adopted; if estimates shift downward, the lower values will be adopted. Now the earlier inputs are planned to create a favorable setting for later inputs, in the knowledge that the later inputs are uncertain. It is impossible to judge the appropriateness of planned early inputs without knowing what possible courses the later inputs may follow if a given series of early inputs is adopted. The problem of fixing the early inputs is thus different (and, of course, more simple) if the possible ultimate distributions of selling-price probabilities have similar expectation values

[12] As Lindahl puts it (*op. cit.*, p. 44), "There exists a *mutual* interconnection between the present and future actions included in the plan."

than if the likelihood function of their expectation values has a wide dispersion.

8. The writer has not been able to satisfy himself that the relative dispersion of profit expectations must necessarily either rise or fall as flexibility is introduced. It is readily demonstrated that cases *may* arise under which the standard deviation (and a fortiori the coefficient of variation) of the profit distribution is reduced by flexibility; and the conditions governing these cases suggest that this is the normal situation. But the writer has found it possible to concoct special cases under which flexibility raises the standard deviation; and while the downward bias of the coefficient of variation resulting from the rising expectation prevents a rise of relative dispersion in these particular cases, it seems impossible to prove in general that the coefficient of variation cannot rise.

9. In any event, the mere fact that the expectation value is raised by embracing flexibility is enough to overthrow the standard assumption that measures for meeting uncertainty necessarily involve sacrificing part of the income expectation in order to gain security. The more important devices for meeting uncertainty—maintenance of cash balances and inventories, selection of unspecialized equipment, choice of processes under which intermediate products may be shifted to different dates or types of output, etc.—contribute to flexibility and tend to raise income expectations.[13] The device of long-term contracts (since it makes plans rigid) is not quite on the same footing; but it is not plain a priori that it tends to lower expectations.[14]

The economist's instinct that analysis of uncertainty requires consideration of the higher moments of the estimate distributions is sound. But the assumption of "risk aversion" which is ordinarily made at the beginning of the discussion is unnecessary in a first approximation. The fact that the higher moments of the estimate distributions, under flexibility, enter into the expectation value of the profit distributions means that the central problems of uncertainty can be posed and largely solved under the assumptions of "risk neutrality."

[13] At least they raise expectation of income if their own costs are zero. If their costs are high enough (if, e.g., unspecialized equipment which will do given work is more expensive than specialized equipment with like productiveness and operating costs), the *net* effect on income expectations can become unfavorable.

[14] Insurance is also in a special category. It involves rigidity of a certain money-outlay schedule; but in return it guarantees that certain options will not be closed in case of loss, so that it contributes to operating flexibility. Its relation to uncertainty (like that of "liquid" assets) is deeply involved in the problem of *capital rationing*, which cannot be discussed here for lack of space. The writer has gone into the problem in some detail in *Anticipations, Uncertainty, and Dynamic Planning*, pp. 39 ff.

Dropping this assumption, whose use greatly simplifies the early stages of reasoning, leads only to secondary qualifications.[15]

APPENDIX

A-1. A general mathematical demonstration of the principles of the text is beyond the scope of this note. But this paper is aimed to controvert two opinions: (*a*) that devices for meeting uncertainty lower profit expectations and (*b*) that to find a theoretical role for the higher moments of estimate distributions we must suppose the planner to be concerned about the higher moments of the profit distribution. Being essentially negative, these opinions can be overthrown by showing even one special case to the contrary. Such a case will be analyzed in this appendix.

A-2. Suppose a producer is planning to produce a commodity X, to be sold two intervals of time hence, at date t_2. He plans to use an input A applied immediately and an input B to be applied at an intermediate date t_1. Prices of the inputs are considered certain; and, by adopting suitable units, we can make both prices unity. The price of output is uncertain: at t_0 the producer recognizes n possible prices, $P_1, P_2, P_3, \cdots, P_n$, with likelihoods, respectively, of $K_1, K_2, K_3, \cdots, K_n$. He expects that before date t_1 some one of these prices (he does not know which) will become certain. Input and output are bound together by a production function $X = F(A, B)$.

A-3. To begin with, we may assume that institutional pressure makes it necessary to contract in advance for both inputs. In this case, if the price is P_j, the profit will be $N_j = P_j \cdot F(A, B) - A - B$. If the producer wishes to maximize his *expectation of profit*, we have:

$$E(N) = \sum_{j=1}^{j=n} K_j[P_j \cdot F(A, B) - A - B] \tag{1}$$

$$= \sum_{j=1}^{j=n} [K_j P_j] F(A, B) - A - B = \text{maximum}.$$

In short, if we call the *price expectation*, $\sum K_j P_j$, $E(P)$, he should behave as though a price of $E(P)$ were certain. The solution, of course, is found by setting the partial derivatives equal to zero. Then:

$$\sum_{j=1}^{j=n} K_j P_j \cdot \frac{\partial F(A, B)}{\partial A} = 1 \tag{2}$$

$$\sum_{j=1}^{j=n} K_j P_j \cdot \frac{\partial F(A, B)}{\partial B} = 1 .$$

Equations (2) will yield optimum solutions for the inputs A and B and by implication for X; and given A, B, X, and $E(P)$ the profit expectation is determined.

[15] Recognition of the flexibility problem, however, greatly changes the complexion of the conclusions reached by some students of the problem. Since increasing flexibility is likely both to raise the profit expectation and to lower the profit dispersion, a planner with either a neutral attitude or a distaste for danger cannot forego it. If devices of this sort are neglected, it will be by a planner *with a positive liking for danger*—a finding counter to what seems to be the general view.

No attribute of the price-distribution function except its average $E(P)$ has any effect on the outcome.

A-4. Suppose now that the producer is set free to postpone decision on the input B until the price of X has been ascertained. We may think of him as fixing input A provisionally at a level A_m. He is still free, however, to vary input B and output X. Now for a price P_j, his profit expectation is

$$N_{m,j} = P_j \cdot F(A_m, B) - A_m - B\,.$$

This may be influenced by his choice of B; plainly if he wants the highest expectation (given the price, his initial decision on A, and the production function), he will set

$$N_{m,j} = P_j \cdot F(A_m, B) - A_m - B = \mathit{maximum}\,. \tag{3}$$

From this, by differentiation,

$$P_j(\partial F/\partial B) = 1\,. \tag{4}$$

This gives a solution for the optimum value of B, which we may designate as B^*. Making B^* explicit, we have

$$B^* = G(P_j, A_m)\,. \tag{5}$$

On general economic grounds we may suppose the partial derivatives of G with respect to both P_j and A_m to be positive.

Having determined B^*, we have by implication determined both the optimum output (X^*) and the corresponding profit (N^*). Both these magnitudes will be increasing functions of P_j and A_m. For a given A_m, we may summate over all possible prices, which gives us

$$E_m(N) = \sum_{j=1}^{j=n} K_j N^*_{m,j} = \sum_{j=1}^{j=n} K_j X^*(P_j, A_m) - A_m - \sum^{j=n} K_j B^*\,. \tag{6}$$

If the maximum expectation is desired, A_m should be so chosen as to maximize this expression. Setting the partial derivative of $E_m(N)$ with respect to A_m equal to zero, we obtain:

$$\sum_{j=1}^{j=n} K_j(\partial X^*/\partial A_m) - 1 - \sum_{j=1}^{j=n} K_j(\partial B^*/\partial A) = 0\,. \tag{7}$$

Solving this equation will yield an optimum value for A_m, which will depend on the dispersion as well as the expectation value of the price distribution.

A-5. If we express the expectation values of profits in terms of our other magnitudes, we find the profit expectation, in the first case (both inputs determined simultaneously), to be

$$E_0(N) = E(P)X - A - B\,, \tag{8}$$

and, in the second case (inputs determined successively),

$$E_m(N) = E(PX^*) - A_m - E(B^*)\,. \tag{9}$$

It is readily demonstrated that E_m exceeds E_0. For we might have laid out the flexible plan subject to the restriction that A_m be fixed equal to the equilibrium A of the rigid plan, and that the weighted sum of contemplated possible B's be

fixed equal to the equilibrium B of the rigid plan. In this case the last two terms of the expectations would be identical. The expectation of X^* might be somewhat smaller than the equilibrium value of X under the rigid plan if some of the values of B^* greatly exceeded B (owing to the action of diminishing returns to B on the fixed A_m); but by suitably balancing values of B^* in excess of B against values less than B, a plan could be found under which the expectation of X^* would fall short of X by less than any assigned quantity. Since the values of X will be positively correlated with the prices, $E(PX^*)$ must exceed $E(P)X$ when the expectations of X are approximately equal. A fortiori, if the restrictions are removed, the flexible plan is even more clearly the better.

A-6. The fact that higher moments of the price distribution will enter into the expectation of profits under flexibility is plain in the light of the determination of X^*. Obviously, given A_m, the optimum values B^* and X^* must be positively correlated with price. This implies that the function determining X^* must contain a term or terms involving the price with both coefficient and exponent of the same sign. If both coefficient and exponent are positive (as they will be if we adopt any plausible production function), then when X^* is multiplied by the price in obtaining the profit expectation the resulting expression will contain some power of P with an exponent greater than unity.[1]

A-7. The relation between the higher moments of the profit distributions, as was mentioned in the text, is complex. For the case of simultaneous determination, the variance about the expectation is simply X_0^2 times the second moment of the price distribution, when X_0 is the output offering the maximum expectation. The variance with successive determination includes the second moment of PX^*; and by the argument of the previous paragraph it will therefore contain moments of P beyond (or short of) the second. The writer has been unable to discover any proof that the coefficient of variation is necessarily less with flexibility, though he has been equally unable to devise any numerical special case under which it is greater.

A-8. Finally, we may consider the effects of substituting probability distributions with expectation values of $P_1, P_2, P_3, \cdots, P_n$ for the supposed certain price estimates anticipated for t_1. So long as it is impossible to defer decision on the input B beyond t_1, there is no effect on the optimum policy either under simultaneous or under successive determination. Furthermore, the expectation of profits under either system of planning will be unaffected. But the variance both of prices and of profits will be increased under both methods, and in the same proportion. Accordingly the fact that estimates are not single valued at the date of final decision will affect planning only if a subsidiary goal of planning is to hold down the profit dispersion.

[1] If the expression for X^* contains a negative term with price in the denominator (which will also give a positive correlation of X^* and P_j), then X^*P_j will contain P to some fractional power or even negative power. But in any event moments other than the first will figure in the profit expectation.

INTERNATIONAL TRANSFERS AND THE TERMS OF TRADE: AN EXTENSION OF PIGOU'S ANALYSIS

M. BRONFENBRENNER
Division of Research and Statistics, United States Treasury

INTRODUCTION

TWO countries, X and Y, trade with each other freely under static conditions — meaning particularly under unchanging utility functions and production techniques. Country X ships "X-goods" (symbolized simply by X) to Y, and to Y alone. Reciprocally, Country Y ships "Y-goods" (symbolized by Y) to X, and to X alone. There are no third countries in this system; Country X is taken to represent "the rest of the world." The $1/n$ inhabitants of X are identical; so also are the $1/m$ inhabitants of Y. To the native of X, his country's export commodity has a final utility f, and imports from Y have a final utility ϕ. To the native of Y, the final utility of X-goods is F, and that of Y-goods is ψ.[1]

When trade between the two countries is in equilibrium, this Jevonian relation holds for every inhabitant of X and Y:[2]

$$\frac{\phi(nY)}{f(nX)}=\frac{X}{Y}=\frac{\psi(mY)}{F(mX)} \quad \text{or} \quad \frac{\phi_1}{f_1}=\frac{X_1}{Y_1}=\frac{\psi_1}{F_1}. \tag{1}$$

Suppose Country Y to undertake to transfer R units per period to Country X, R being measured in Y-goods. In Pigou's example the transfer constitutes the payment of a war indemnity, German reparations. It may equally well constitute the making or the repayment of any international loan.

The "gross barter" terms of trade are measured by (X/Y), the price of Y-goods in terms of X-goods. In case the terms of trade turn against the paying country, Y, (i.e., in case the ratio X/Y increases in size), Y must assume not only the initial burden of raising the payment to be transferred but also an additional "transfer burden" aris-

[1] This notation and these assumptions are those of A. C. Pigou's essay, "The Effect of Reparations on the Ratio of International Interchange," *Economic Journal*, XLII (December, 1932), 532–43.

[2] This condition is presented as necessary to, not sufficient for, equilibrium. The commodities X and Y are measured in physical units. The quantities nX, nY, $\cdots$, are quantities exported or imported per individual per period. Thus, $nX = X/(1/n)$, etc.

ing in consequence of the unfavorable shift in the terms of trade. If the terms of trade do not change as a result of the transfer, there is no such transfer burden. If the shift in the terms of trade is in favor of Y (if X/Y decreases in size), the effect of transfer is to lighten the original burden of raising the funds to be transferred. A type of "transfer bounty" is involved.

In the essay previously cited Pigou attempts to determine the conditions governing the direction of shift of the terms of trade and also the probability of a shift in any particular direction. He has shown, without requiring additional assumptions, that the terms of trade would turn in favor of the receiving Country X, remain unchanged, or turn in favor of the paying country Y, according as

$$\frac{\phi'}{f'} \lesseqgtr \frac{\psi'}{F'}, \tag{2}$$

the primes denoting first derivatives with respect to Y in the numerators and with respect to X in the denominators. In translating inequalities (2) into the language of price and elasticity, Pigou introduced further postulates, namely, linearity of all utility functions and "similarity" of the utility functions of the inhabitants of the two countries.[3] In this paper I shall attempt the translation, using what seems to me a less extreme set of additional postulates. The final result will be somewhat more complex than Pigou's and will take formal account of more of the relevant factors. Unfortunately, it would appear that the effect of the additional terms is difficult to determine a priori and that only the results of the Pigou analysis are of immediate applicability to the analysis of practical problems.

EXPANSION OF THE PIGOVIAN INEQUALITY

From the general theory of exchange, we have, for the representative inhabitant of X and Y, respectively,

$$\mu_x = \frac{\phi}{p_y} = \frac{f}{p_x} = \cdots \quad \text{and} \quad \mu_y = \frac{\psi}{\pi_y} = \frac{F}{\pi_x} = \cdots .$$

The common ratios, μ_x and μ_y, are the two final utilities of money income. The terms in p are prices in X-money; those in π are prices in Y-money.

When we substitute these results in the fundamental inequalities

[3] For a graphic interpretation of the meaning of Pigou's assumption of "similarity," see Jacob Viner's *Studies in the Theory of International Trade* (New York, 1937), Figs. 3–6, p. 341.

(2), we obtain at once:

$$\frac{\dfrac{d(\mu_x p_y)}{dY}}{\dfrac{d(\mu_x p_x)}{dX}} \lesseqgtr \frac{\dfrac{d(\mu_y \pi_y)}{dY}}{\dfrac{d(\mu_y \pi_x)}{dX}}.$$

Denote by i_x the income of an inhabitant of X, income being defined as simply consumption or $n(xp^0_x + Yp^0_y + \cdots)$. Similarly, denote by i_y the income of an inhabitant of Y, or $m(X\pi^0_x + y\pi^0_y + \cdots)$.[4]

Let us also assume that $\mu_x i_x = k_x$ and $\mu_y i_y = k_y$, by the Bernoullian hypothesis.[5] Since any choice of units of utility is arbitrary, we choose units of μ_x so that k_x is unity. The utilities of the representative natives of X and Y are not comparable, so that we may repeat this procedure with μ_y, setting k_y equal to unity as well, without becoming involved in contradictions.

When we make the indicated substitutions of $(1/i)$ for μ above, and then carry out the differentiations, the result is:

$$\frac{\dfrac{d\,p_y}{i_x dY} - \dfrac{p_y\, di_x}{i^2_x\, dY}}{\dfrac{d\,p_x}{i_x dX} - \dfrac{p_x\, di_x}{i^2_x\, dX}} \lesseqgtr \frac{\dfrac{d\,\pi_y}{i_y dY} - \dfrac{\pi_y\, di_y}{i^2_y\, dY}}{\dfrac{d\,\pi_x}{i_y dX} - \dfrac{\pi_x\, di_y}{i^2_y\, dX}}.$$

This expression may be re-written in terms of export and import price and income flexibilities. The first step in doing this is to expand the total derivatives in terms of partial derivatives:

$$\frac{d\,p_y}{d\,Y} = \frac{\partial\, p_y}{\partial\, Y} + \left[\frac{\partial\, p_y}{\partial X}\frac{d\,X}{d\,Y} + \frac{\partial\, p_y}{\partial\, Z}\frac{d\,Z}{d\,Y} + \cdots + \frac{\partial\, p_y}{\partial\, i_x}\frac{d\,i_x}{d\,Y}\right]$$

. . . .

$$\frac{d\,i_x}{d\,Y} = \frac{\partial\, i_x}{\partial\, Y} + \frac{\partial\, i_x}{\partial\, X}\frac{d\,X}{d\,Y} + \left[\frac{\partial\, i_x}{\partial\, Z}\frac{d\,Z}{d\,Y} + \cdots\right]$$

. . . .

Each of the quantities in brackets is of uncertain sign. The sum

[4] Here the p^0 and π^0 represent "base year" or " pretransfer" prices. We use x and y instead of X and Y, where the domestic consumption of the goods in question is relevant rather than the quantity entering international trade.

[5] Or, more generally, we may assume that the flexibility of the final utility of income is approximately constant in the two countries:

$$\mu_x^{H_x} i_x = k_x \quad \text{and} \quad \mu_y^{H_y} i_y = k_y.$$

The terms in H, which represent the flexibilities, makes little difference to the final solution (see below, n. 7). We omit them for the sake of simplicity.

of the quantities in any given set of brackets is also of uncertain sign. In any case, these sums are probably small in comparison with the other terms. We may neglect these sums and confine our attention to the other terms, at least as a first approximation. Further refinements may do well to consider at least the first and last terms in the upper bracket, even though reactions on domestic commodities (Z) be neglected.

Denote by Γ_y the import price flexibility of Y. When commodity Y is imported,

$$\Gamma_y = \frac{\partial p_y}{\partial Y}\frac{Y}{p_y}.$$

When this commodity is exported, the same mathematical expression[6] defines the export price flexibility of Y and is symbolized by G_y.

Denote by Γ'_y the import income flexibility of Y. When commodity Y is imported,

$$\Gamma'_y = \frac{\partial i}{\partial Y}\frac{Y}{i}.$$

When this commodity is exported, the same mathematical expression defines the export income flexibility of Y and is symbolized by G'_y.

In terms of these flexibilities, the total derivatives of the last set of inequalities may be written

$$\frac{d\,p_y}{d\,Y} = \Gamma_y \frac{p_y}{Y} \qquad , \quad \frac{d\,p_x}{d\,X} = G_x \frac{p_x}{X}\ , \cdots$$

$$\frac{d\,i_x}{d\,Y} = \Gamma'_y \frac{i_x}{Y} + G'_x \frac{i_x}{X}\frac{d\,X}{d\,Y}\ , \frac{d\,i_x}{d\,X} = G'_x \frac{i_x}{X} + \Gamma'_y \frac{i_x}{Y}\frac{d\,Y}{d\,X}\ , \cdots ,$$

whereupon the inequalities themselves become

$$\frac{\dfrac{\Gamma_y p_y}{Y\,i_x} - \dfrac{\Gamma'_y p_y}{Y\,i_x} - \dfrac{G'_x p_y}{X\,i_x}\dfrac{d\,X}{d\,Y}}{\dfrac{G_x p_x}{X\,i_x} - \dfrac{G'_x p_x}{X\,i_x} - \dfrac{\Gamma'_y p_x}{Y\,i_x}\dfrac{d\,Y}{d\,X}} \lessgtr \frac{\dfrac{G_y \pi_y}{Y\,i_y} - \dfrac{G'_y \pi_y}{Y\,i_y} - \dfrac{\Gamma'_x \pi_y}{X\,i_y}\dfrac{d\,X}{d\,Y}}{\dfrac{\Gamma_x \pi_x}{X\,i_y} - \dfrac{\Gamma'_x \pi_x}{X\,i_y} - \dfrac{G'_y \pi_x}{Y\,i_y}\dfrac{d\,Y}{d\,X}}$$

This expression can be simplified considerably. The terms in income cancel at once. Since X and Y are internationally traded commodities, and since we assume free trade conditions, the ratio of their prices is approximately equal in the two countries. We may therefore

[6] With a change from p to π, when necessary.

multiply the left side by (Yp_x/Xp_y), and the right side by $(Y\pi_x/X\pi_y)$. This leaves

$$\frac{\Gamma_y - \Gamma'_y - G'_x \dfrac{(YdX)}{(XdY)}}{G_x - G'_x - \Gamma'_y \dfrac{(XdY)}{(YdX)}} \lesseqgtr \frac{G_y - G'_y - \Gamma'_x \dfrac{(YdX)}{(XdY)}}{\Gamma_x - \Gamma'_x - G'_y \dfrac{(XdY)}{(YdX)}}.$$

The third terms of both numerators and denominators are interrelation factors. They include terms in (YdX/XdY), which might be christened intercommodity elasticities. We shall call these interrelation factors c when used with G', and γ when used with Γ'. Each interrelation factor will be given the subscript of the corresponding income flexibility. The sign of the interrelation factor will be opposite to that of the corresponding income flexibility, however, since the effect of the transfer is to render dX negative, dY positive, and (dY/dX) negative. This substitution gives us

$$\frac{\Gamma_y - \Gamma'_y - c_x}{G_x - G'_x - \gamma_y} \lesseqgtr \frac{G_y - G'_y - \gamma_x}{\Gamma_x - \Gamma'_x - c_y}.$$

The signs of the various terms are normally as follows:

Positive	Negative
G	Γ
Γ'	G'
c	γ

The left-hand denominator of the last expression is therefore positive, the right-hand denominator negative. When, in simplifying this expression, we multiply it by the product of the denominators, we must change the direction of the inequalities, and obtain as the result:

$$\begin{aligned}&\Gamma_x\Gamma_y + \Gamma'_x\Gamma'_y + c_x c_y + \sum_{i\neq j}\Gamma'_i c_j - \Big(\sum_{i\neq j}\Gamma_i\Gamma'_j + \sum_{i\neq j}\Gamma_i c_j\Big) \gtreqless \\ &G_x G_y + G'_x G'_y + \gamma_x\gamma_y + \sum_{i\neq j} G'_i\gamma_j - \Big(\sum_{i\neq j} G_i G'_j + \sum_{i\neq j} G_i\gamma_j\Big).\end{aligned}$$

This can be simplified further, since we can easily see that

$$c_x c_y = G'_x G'_y \quad \text{and} \quad \gamma_x\gamma_y = \Gamma'_x\Gamma'_y.$$

Subtracting out equal terms, we obtain finally:[7]

[7] These inequalities are slightly more complex when we permit the flexibility of the final utility of money income to depart from unity. The formula corresponding to (3) is then:

$$\Gamma_x \Gamma_y + \sum_{i \neq j} \Gamma'_i c_j - \left(\sum_{i \neq j} \Gamma_i \Gamma'_j + \sum_{i \neq j} \Gamma_i c_j\right) \gtrless G_x G_y + \sum_{i \neq j} G'_i \gamma_j - \left(\sum_{i \neq j} G_i G'_j + \sum_{i \neq j} G_i \gamma_j\right). \tag{3}$$

When all income elasticities and interrelation factors are neglected, we have as a special case of equation (3) Pigou's mathematical condition, which was also derived by Yntema.[8]

$$\Gamma_x \Gamma_y \gtrless G_x G_y \text{, or, in terms of elasticities, } e_x e_y \gtrless \eta_x \eta_y .$$

Inequalities (3) have the same economic meaning regarding terms of trade as do the fundamental inequalities (2), from which they were derived. If the function of demand flexibilities on the left exceeds the similar function of supply flexibilities on the right, the terms of trade move in favor of X, the receiving country, and there is a transfer burden. If the reverse is true, the movement of the terms of trade favors Y, the paying country, and there is a transfer bounty. If the two functions are equal, the terms of trade are unchanged, and there is neither burden nor bounty.

THE PROBABILITIES

Pigou has stated his reasons for believing the upper branch of (3), the shift in favor of the receiving country and involving transfer burden, to be the most probable alternative in practice:

> Since the receiving country spends a large amount of its income in securing its own goods and a small part in securing the paying country's goods, there is a general presumption that, for a given absolute change in the quantity of produc-

---

$$\Gamma_x \Gamma_y + \frac{1}{H_x H_y} \left(\sum_{i \neq j} \Gamma'_i c_j\right) - \left(\sum_{i \neq j} \frac{\Gamma_i \Gamma'_j}{H_j} + \sum_{i \neq j} \frac{\Gamma_i c_j}{H_j}\right) \gtrless G_x G_y + \frac{1}{H_x H_y} \left(\sum_{i \neq j} G'_i \gamma_j\right) - \left(\sum_{i \neq j} \frac{G_i G'_j}{H_j} + \sum_{i \neq j} \frac{G_i \gamma_j}{H_j}\right). \tag{3a}$$

Since terms in H are present symmetrically on both sides, it is very unlikely that their presence will disturb any qualitative conclusions we draw from the simpler form (3), even though H_x may differ materially from H_y.

It is frequently argued, in quite another connection—the theoretical justification of tax progression—that the marginal utility of income is inelastic. If this is correct, the marginal utility flexibilities H_i are probably greater than unity. This would tend to minimize the quantitative importance of all terms in (3*a*) which are not found in Pigou's solution.

[8] For T. O. Yntema's determinantal derivation see chap. v (esp. pp. 89–92) of his *Mathematical Reformulation of the Theory of International Trade* (Chicago, 1931).

tive force devoted to its own goods, and so in the quantity of its goods, the change in marginal utility will be much smaller than the change in marginal utility that will result from an equal absolute change in the quantity of its productive force devoted to the paying country's goods, and the quantity of those goods.

This implies that f' is (numerically) much smaller than ϕ', and so that ϕ'/f' is (numerically) much larger than unity. By analogous reasoning ψ' is likely to be smaller than F'. Therefore ψ'/F' is likely to be numerically smaller than unity. Thus in general there appears to be a very high probability that ϕ'/f' is numerically much larger than ψ'/F'; i. e., that $e_x\, e_y > \eta_x\, \eta_y$, which is the condition that the real ratio of interchange is turned against the paying country.[9]

This is a realistic statement of the case. Objection has been made to it on the ground that its very realism is due to departures from Pigou's assumptions.[10] If paying and receiving countries are of equal economic importance, the only reason to suppose that each country consumes less imported than domestic goods (which supposition forms the basis of Pigou's conclusion) is the existence of the very domestic commodities, customs duties, transport costs, etc., from which abstraction is apparently being made. And if one country is larger than the other, particularly if it be "the rest of the world," one should expect f' and F' to each be larger or smaller than ϕ' and ψ', respectively, in the absence of barriers. The results would then be quite indeterminate as regards the relative probabilities of the branches of (3).

As an alternative method of ascertaining the relative probabilities, we might resort to elementary observations of the comparative keenness of buyers and sellers on any market for consumption goods. It requires generally a greater price change to cause appreciable alterations in the amounts ultimate consumers will buy, whereas minute variations in existing or expected prices suffice to alter producers' outputs and supplies. In general, suppliers (including exporters with a choice between several markets for disposal of their products) are more sensitive to differential profits than are consumers to differential satisfactions. *Ceteris paribus,* then, $|e_x| > |\eta_x|$,[11] and so, in gen-

[9] *Op. cit.*, pp. 539 f. (I have taken the liberty of substituting certain of my symbols for certain of Pigou's and also of referring to "the paying country" and "the receiving country" instead of his "Germany" and "England.")

[10] Notably by Viner, *op. cit.*, pp. 340 f.

[11] Some such line of reasoning as this may underlie Yntema's bald statement (*op. cit.*, p. 92) that the average elasticity of supply is numerically greater than the average elasticity of demand, and therefore that the most probable branch of (3) favors the receiving country.

Another factor tending to support Yntema's statement is that, under competitive conditions, changes in the number of plants tend to make supply elasticities approach infinity in the long run, at the point of minimum average cost, in the absence of specialization of resources, while no such tendency exists on the demand side. The relevance of long-run considerations to this essentially short-run problem is, however, dubious.

eral, terms of trade may be most likely to shift in favor of receiving countries.

This contention is invalid when the purchase as well as (or instead of) the sale is carried on for profit instead of final consumption. Many of the purchases in international trade are of this type, since many of the goods which move are raw materials, intermediate products, and capital goods. Many of the consumers' goods which move are also shipped from country to country on the orders of dealers bent on profit.

It is difficult to draw conclusions from the income flexibilities and interrelation factors of (3) because of their tendency toward mutual offsetting. To illustrate: Suppose Γ'_x to be (numerically) high, which condition would tend to increase the probability of movement in favor of Y. But in that case γ_x (which equals $\Gamma'_x\, YdX/XdY$) would also be (numerically) large, which condition would increase the probability of a shift in favor of X. The net result depends on the intercommodity elasticities of the particular case. It defies unequivocal determination a priori.

Two subsidiary problems may arise for discussion at this point: (i) Does the numerical size of the inequality (3) indicate the approximate extent of the shift in terms of trade required to effect the transfer? and (ii) under what conditions is transfer not only burdensome but completely impossible?

(i) The greater the numerical value of the inequality (3), the greater the transfer burden (or bounty, depending on the direction of the inequality), provided that all significant terms of (3) remain constant as the transfer gets under way and the terms of trade begin their shift. Normally, however, we cannot expect such constancy over any considerable range.

Assume the existence of a transfer burden. Then, as incomes in the receiving country increase, the relative cheapness of the paying country's goods does likewise. These changes combine to bring about a rise in the receiving country's import demand flexibility. Their converses in the paying country offset them to an indeterminate extent, as regards their effects on (3), since the paying country's import demand flexibility should fall as incomes fall and foreign goods become relatively more expensive. These income and price changes should also induce a fall in the export supply flexibility in the paying country, offset as regards effects on (3) by a rise in the export supply flexibility in the receiving country.

In case a transfer bounty exists, it is difficult to forecast the direction of any of the shifts in flexibility, since the income and rela-

tive price changes all tend to offset each other.

It is therefore quite unsafe not only to hazard more than the merest guess as to the extent of transfer burden or bounty from the numerical value of the inequality (3) but also to state the direction of bias involved in the first guess.

(ii) The economic impossibility of an international transfer would be indicated by so extreme a shift in the terms of trade when such a transfer was undertaken that these terms would approach zero or infinity or become indeterminate. Inspection of Pigou's original Jevonian relation (1) will show that the first two cases imply that one or the other country's export commodity becomes a free good, while the third is impossible unless the marginal utilities of the two countries' goods become simultaneously zero or infinite. None of these cases, it should be clear, has any practical significance, except under a complete embargo. Transfer is never economically impossible, though the burdens involved may be and have been so great as to render it politically so.

NECESSITIES AND LUXURIES

So much for the general problem. Let us now consider certain special conditions which may tend to reinforce or contradict our general conclusion that shifts in terms of trade will probably favor receiving countries when transfers occur. Many of our results will not be original, having been reached by previous writers in arithmetical examples or on purely theoretical grounds.[12]

If either country's export commodity is a necessity or is complementary with domestic commodities,[13] it will probably have a higher export supply price flexibility than otherwise. Conversely, if either country's export commodity is a luxury, or is competitive in consumption with domestic commodities, this flexibility will tend to be low. Therefore, we may conclude that when either country's ex-

[12] A good survey of the literature dealing with the transfer problem under various conditions is given by Viner, *op. cit.*, pp. 326–38. Gottfried von Haberler's *Theory of International Trade* (English trans.; New York, 1936) also deals extensively with the literature on these problems in chap. vii. The first-mentioned work stresses the English-language literature; the second, the contributions in the German language.

[13] By "necessity" I mean, following a generally accepted definition, a good on which is spent a steadily smaller percentage of a representative consumer's income as that income increases, and by a "luxury" a good for which the reverse is true. It is quite irrelevant to our problem here which of the several conflicting definitions of complementarity is adopted, though the writer's personal predilections favor that set forth by J. R. Hicks, *Value and Capital* (Oxford, 1939), pp. 44 and 311.

port commodity has the status of a necessity, or is highly complementary with domestic commodities, terms of trade are rendered especially likely to shift in favor of paying countries. The opposite conditions increase the likelihood of shifts in favor of receiving countries.

If either country's import is a necessity or is complementary with home commodities in consumption, it will have a higher import demand price flexibility than otherwise, and vice versa. The effects here are opposite to those in the preceding case. If either country's import is a necessity, the terms of trade are rendered more likely to turn in favor of the receiving country when transfers occur. It is the paying country which benefits from a condition in which either country's import is a luxury or is competitive with domestic commodities in consumption.

It should be noted (i) that our conclusions are perfectly symmetrical as between countries; that (ii) if both countries' imports (or exports) are necessities (or luxuries), these conclusions tend to be reinforced, not offset; and that (iii) consumption, for purposes of this discussion, is inclusive of production of other goods.

In the actual world a larger proportion both of a country's imports and of its exports is likely to be made up of luxuries than is the country's domestic production or consumption. The effects of these two tendencies tend to nullify each other, so that we cannot draw definite conclusions as to the effect of these tendencies upon the direction of inequality (3). When some country is a necessity exporter and luxury importer, like the United States or the Australian Commonwealth, its peculiar status will tend to bring about shifts in the terms of trade in favor of paying countries in international transfers in which the United States or Australia play principal roles.

INFERIOR COMMODITIES

We have avoided complications involved by the presence of inferior commodities—commodities with negative domestic income elasticities—in international trade. Their effects generally seem to be formally offsetting, and arithmetically determinate only in special cases.

If either country's export commodity is considered inferior, then a G', on the right side of (3) will be small. The corresponding c will, at the same time, become small on the left side of this inequality, and the net conclusion is arithmetically uncertain. If either country considers its imported commodity inferior, Γ' will be small on the left side of (3), γ correspondingly small on the right side, and the net

conclusion will again be doubtful.[14] The case is again one of symmetry as between countries, and repetitions of phenomena have reinforcing rather than offsetting force.

Inferiorities of themselves do not seem generally to influence the direction of shifts in terms of trade under international transfers, though it is quite possible to construct arithmetical examples in which their influence is marked.[15] However, if inferior imports have low demand price flexibilities (as they well may), and inferior exports have low supply price flexibilities, we may say that an inferior import increases the probability of a shift in favor of the paying country, and an inferior export increases the probabiltiy of a shift in the opposite direction.

A highly special case arises when one commodity is considered inferior in both countries, if the extent of inferiority is very marked. Suppose that the net effect of the transfer is to cause this commodity (whether Y or X makes no difference) to be consumed in larger quantity than before in the paying country, and in smaller quantity than before in the receiving country. When this is the case, (dX/dY) which we have assumed negative, becomes positive, and even inequalities (2) no longer hold, as can be seen by one attempting to retrace Pigou's analysis.

TARIFFS AND BOUNTIES

The major determinate effect of a revenue tariff imposed by either country on the other country's export commodity is probably to lower its own Γ for the foreign commodity. This can be shown most easily by reference to the definition of the demand price flexibility. The effect of the tariff is to raise the price the imposing country pays for its imports and lower the quantity of those imports. There may be a partially (though probably not completely) offsetting increase in $\partial p/\partial q$ as the price rises. On balance, however, any revenue duty imposed by either country lowers a Γ, therefore lowers

[14] It may occur to the reader that when a country's export commodity is inferior, G' will become positive instead of negative, or that, when it is the import commodity which is inferior, Γ' will become negative instead of positive. The corresponding income elasticities do in fact change sign when commodities become inferior. Since we are dealing with *partial* derivatives, however, we cannot consider an income flexibility as the reciprocal of an income elasticity, identical with it as regards sign. For whatever be the sign of (say) $\partial X/\partial i$, $\partial i/\partial X$ is always positive when the commodity is imported and negative when it is exported, unless the commodity is a positive nuisance, and the sign of this derivative determines the signs of G' and Γ'.

[15] See, e.g., the example selected and worked out by Wassily Leontief in his essay, "Note on the Pure Theory of Capital Transfer," included in *Explorations in Economics* (New York, 1936), pp. 84–92.

the left side of (3), and tends to turn the terms of trade in favor of the paying country. The possible serviceability of this argument as justifying high revenue tariffs in debtor countries should be clear.

If the duty is protective, giving rise to increased domestic supplies of the commodity formerly imported, the provision of substitutes will still further decrease Γ in the imposing country and tend still further to shift terms of trade in favor of paying countries when transfers occur.[16]

The effects of bounties imposed by either country are somewhat more complex. A *production* bounty's major effect is a change (probably an increase) in the foreign country's import demand price flexibility. The price the foreigner pays is lowered, the quantity obtained generally increased, probably without very marked offsetting effects on the slope of his demand curve. This increase in demand flexibility tends to increase the likelihood of a shift in the terms of trade favorably to receiving countries in case of transfers.

An export bounty also has an effect on the imposing country's export supply price flexibility. The bounty tends to lower the export price of the commodity, and increase the quantity exported at any given price. As a result, it raises G, and the effect of the higher G tends to counteract that of the higher Γ of the other country.[17] The net effect on the probabilities of shift in the terms of trade depends on the numerical values in specific cases.

The case of ordinary dumping can be subsumed under the general head of export bounty. The usually implicit nature of the "bounty" involved and its payment by nonfiscal means do not affect the foregoing discussion.

The cases dealt with in this section differ from those discussed previously in that they are not perfectly symmetrical as between countries. For, in deriving (3), we assumed p_y/p_x approximately equal to π_y/π_x. When tariffs or export bounties are in operation, or when transport costs are high, this is no longer permissible, and our results must be modified accordingly.

When a tariff is imposed by X (the receiving country), $p_y/p_x >$

[16] The conflict of these results with those reached by G. A. Elliott in his note entitled "Protective Tariffs, Tributes, and Terms of Trade," *Journal of Political Economy*, XLV (December, 1937), 304–7 is more apparent than real. Elliott deals with a case in which a *prohibitive* tariff is imposed on one of two export commodities of a paying country. Moreover, he is concerned with "factorial" rather than "gross barter" terms of trade, and it is the "factorial" terms which he shows to be shifted in favor of receiving countries under protection.

[17] The same argument applies, except for the differential between domestic and export prices, when the bounty is a production bounty. It is likely to be less important in that case, however, unless the domestic market for the bounty-fed commodity is a minor portion of its world market.

π_y/π_x. We can see, by reviewing the derivation of (3) that this inequality would tend to increase the probability of a shift in favor of the receiving country, *ceteris paribus*. Whenever the tariff is imposed by the paying country, the relative price effect is the opposite. When an export bounty is paid by a receiving country, or receiving country producers have recourse to dumping, $p_y/p_x > \pi_y/\pi_x$, with increased likelihood of a shift in favor of the receiving country. Conversely, use of export bounties and dumping by paying countries tend to increase the probability of favorable shifts in the terms of trade through their effects on relative prices.

III
STATISTICAL THEORY

ECONOMIC INTERDEPENDENCE AND STATISTICAL ANALYSIS[1]

J. MARSCHAK
New School for Social Research

I. CONDITIONS OF INDUCTIVE STUDY

ECONOMIC equations derived from experience, like Henry Schultz's demand curves for agricultural commodities, have, or strive to attain, this practical importance: they should make it possible to estimate the values which one of the variables, the "predictand" (e.g., demand), will assume when other variables, the "predictors" (e.g., income and price), are made to assume given values.[2] There is, however, a proviso, silently admitted in this as in any other inductive work, whether quantitative or not. If we use the word "magnitude" in a broad sense to include characteristics which can assume the values "presence" or "absence" but which cannot be measured; and if we use the words "observation period" and "prediction period" so as to include observations and predictions not only over time but, for example, within and between geographical areas (or any other samples), then we can formulate the proviso as follows.

Let y be the predictand, let $x_1, \cdots, x_n$ be used by the investigator as predictors, and let $a_1, a_2, \cdots$ denote "all other magnitudes in the world"; some, but not all, of the a's enter the proposed relationship explicitly and are called its parameters. Then every a either must remain constant or its variations must have no significant influence on the value of y. This must be true both during the observation period and during the prediction period.

If the condition is not fulfilled during the observation period, the relationship between y and the x's is a "spurious" one. If the condition is fulfilled during the observation period but not during the prediction period, the case is called "lack of homogeneousness" or of "uniformity" as between the two periods. It is assumed in all inductive work that the investigator tries his best to take over all the relevant

[1] The article could not have been written without the stimulating influence of talks with T. Haavelmo. Acknowledgments are due also to O. Lange, J. Mosak, and A. Wald with each of whom several important points were discussed. The article was written in March, 1940.

[2] The terms "predictand" and "predictors" are H. Hotelling's, I believe.

factors into the "x" group—to use, in other words, a "complete list of factors"—and it is also assumed that the world is not so utterly chaotic and capricious as to make the list of x's infinitely long, even if a comparatively short time (or space, or any other characteristic which may reasonably be connected with some important factors) lies between observation period and prediction period. Those historians who sincerely believe that history does not repeat itself must resign themselves to a selective description of single events. Economists have not, in general, thus resigned themselves. Even where the description of "economic institutions" militantly opposes theory, the existence of "invariants" is assumed, and, in fact, the interest is centered on "standing rules" ("institutions").

II. INTERDEPENDENCE

When a single economic equation—a "demand curve," for example—is used to predict y (demand) for given values of the x's (price, income), the causes determining in turn the x's themselves are not considered: the x's may change because of the weather or because of the government—causes which economists choose to regard as extraneous to their field—or the explanation may be found without recourse to such extraneous fields.[3] In the latter case an equation like Schultz's is regarded as a part of an economic "dynamic system" of equations.

The interdependence between a single empirically found economic equation and the other equations of the system seems, at first sight, both to enhance the importance and to endanger the validity of each single equation. Its importance and interest are increased because economists try to use it not only for the explanation and prediction of the behavior of the "predictand" contained in this equation but also for the explanation of the behavior of all the other variables of the system: for the "Verification of Business Cycles Theories."[4] At the same time, the fact of interdependence causes many workers in the field to subject every single empirical equation to serious doubts: some pitfalls in statistical research on economic data are credited to economic interdependence.

The present article leads, however, to the following conclusion: the single empirical equations are less useful or necessary for the empirical analysis of the system than it is sometimes thought; while, at

[3] Of course, these delimitations need not be hard and fast (see, e.g., Schumpeter, *Business Cycles*, chap. 1).

[4] J. Tinbergen, *Statistical Testing of Business-Cycle Theories* (2 vols.; Geneva, 1939).

the same time, the fact of interdependence is not necessarily fatal to the validity of such single equations for the purpose of prediction.

III. TINBERGEN VERSUS MITCHELL

Tinbergen's verification of business-cycle theories is performed in these steps:

1. A system of economically meaningful and plausible equations (the theory to be tested) is set up. Their number must equal the number of variables (excluding time).
2. For the equations other than identities the numerical values of parameters (such as elasticities, lags, etc.) are estimated either by the regression analysis of time series or from other sources—technological recipes, interviews. In the process some of the theoretically plausible equations are weeded out if contradicted by facts.
3. From the system of equations, with numerical parameters inserted, all variables are eliminated except one—say, x. Thus a functional equation is obtained which explains the value of x at any given point of time by earlier values of x. For example, if linear relationships were assumed throughout, we obtain

$$x(t) = A_0 + A_1x(t-1) + A_2x(t-2) + \cdots. \tag{1}$$

Here x is an implicit function of time. The empirical coefficients A_0, $A_1, \cdots$ are thus obtained in the process of elimination. That is, they are the result of certain operations on those numbers which either the regression analysis or other methods have yielded at stage 2.

4. From the equation (1) the equation of the cycle is derived: x is expressed as an explicit function of time. This is, in the simplest cases, a sum of sine functions (damped, undamped, or antidamped) and of pure exponential functions.
5. The cycles thus derived are compared with the existing time series.

A few important objections, or corrections, can be and have been advanced.

(*a*) A regression coefficient is an estimate only: in economics its error may be very considerable. To obtain the equation (1), regression coefficients obtained in step 2, are operated upon as if they were exact numbers. In principle it should be possible to supplement the operations on the estimates by operations on errors, although the existing formulas for the error of a given function are merely approximations (except in the simplest cases). Would it not be simpler, however, to abstain from regression analysis at the early step 2 and

to postpone the estimation of the A's until step 3? The resulting equation will be similar to (1) but with coefficients A_0, A_1, unknown (although some relationship between them may be known owing to the insertion of the data from "interviews," etc., into the theoretical equations). Similar equations, with coefficients B_0, B_1, $\cdots$, C_0, C_1, can be obtained for the other variables—y, and z. These unknown coefficients can then, in principle, be estimated by statistical methods, although their exact distribution from sample to sample has not yet been found (see below, Sec. VI).

(*b*) More than one equation of the cycle may correspond to a given equation (1); also, conversely, more than one equation (1), and consequently more than one system of theoretical equations may correspond to a given form of the cycle. The former fact has been well known and gave to the economists a somewhat doubtful latitude in choosing their cycles. The latter fact is perhaps less well known.[5] Yet, both do not altogether preclude the possibility of testing hypotheses of the economic theory of business cycles in the light of experience: the observations do restrict the class of viable hypotheses, even if it is impossible to reduce the number of such hypotheses to one.

But the result relevant for our discussion is this: to "explain" the cycle it is unnecessary and, because of errors of regression, even dangerous to compute elasticities, etc., first, and to obtain, by operating on them, the equation of the cycle. It is safer to solve the system of theoretical equations in the abstract and to fill in the numerical data in equations (1) (or into the equations of the cycle) for each separate variate directly from the historical data. Does not this mean that the theoretical system should be directly confronted with Mitchell's empirical cycles for separate variables, omitting the intermediary (Schultz's or similar) single empirical relationships?

These relationships are important when partial decisions must be taken. But they do not help to establish the interdependent system as a whole. Some people think they are rendered meaningless by the existence of the system. To this question we proceed in the next two sections.

IV. VARIATE PARAMETERS

Suppose you fit a curve to a scatter of points representing prices and consumed quantities of a commodity. How do you know whether your curve is the demand or the supply curve? A restatement of an-

[5] Ragnar Frisch, "Statistical versus Theoretical Relations in Economic Macrodynamics" (mimeographed, 1938).

swers given earlier by E. Working (1927),[6] R. Frisch (1933),[7] and others is also given on pages 72–81 of Schultz's *The Theory and Measurement of Demand,* although in a somewhat loose form. The answer is, briefly: the fitted curve approximates closer that of the two theoretical curves which has undergone smaller changes during the observation period. As the simplest example, consider the case of straight lines, subject to parallel shifts only. Write, for the tth observation

$$Y_t = \alpha X_t + U_t; \qquad Y_t = \beta X_t + V_t, \tag{2}$$

where Y_t and X_t are quantity and price, and U_t and v_t measure the varying intercepts of the two curves with the Y-axis, while the slopes α and β remain constant; U and V are variate parameters. Which of the two nonvariate parameters, the slope α or the slope β comes closer to the empirical regression coefficient, say, b_{yx} of Y on X? Solving (2) with respect to X_t and Y_t and subtracting from the respective means, we obtain (lower-case letters denote deviations of X, Y, U, V from their respective means; subscript t dropped for brevity):

$$x = \frac{u - v}{\beta - \alpha}, \qquad y = \frac{u\beta - v\alpha}{\beta - \alpha}$$

$$\therefore b_{yx} = \frac{\sum xy}{\sum x^2} = \frac{\sum [u^2\beta + v^2\alpha - uv(\beta + \alpha)]}{\sum (u^2 + v^2 - 2uv)}$$

$$\therefore \frac{b_{yx} - \alpha}{\beta - b_{yx}} = \frac{\sum u^2 - \sum uv}{\sum v^2 - \sum uv} = \left(\frac{\sigma_u}{\sigma_v} - r_{uv}\right) \Big/ \left(\frac{\sigma_v}{\sigma_u} - r_{uv}\right), \tag{3}$$

where σ_u and σ_v are standard deviations, and r_{uv} is the correlation coefficient, of the "shifts." It is seen from (3) that the regression coefficient is nearer to the slope of the demand than to that of the supply curve, if the former is more "stable" than the latter.[8] (In the case of independent shifts, $r_{uv} = 0$; the differences of the two slopes and the regression coefficient are then in exact proportion to the variances of the respective shifts.) In the limiting case one of the curves (e.g., the first one) is completely fixed: inserting $u = 0$ into (3), we get $b_{yx} = \alpha$ exactly. Graphically, all points of the scatter are inter-

[6] *Quarterly Journal of Economics,* 1927.

[7] *Pitfalls in the Statistical Construction of Demand and Supply Curves* (Leipzig, 1933).

[8] The problem of variate parameters must be clearly distinguished from the problem of "fitting a line when both variables are subject to error" (see A. Wald in *Report of Fifth Annual Research Conference, Cowles Commission* [Chicago, 1939], p. 25; R. G. D. Allen in *Economica,* 1939; see Sec. VI below).

sections of the fixed and the "shifting" curve; consequently, they all lie on the fixed curve. If *both* curves are fixed, $u = v = 0$, and b_{yx} becomes indeterminate. Graphically, if none of the curves is subject to shifts, the "scatter" of their intersection points reduces to one single point; and there is an infinite number of straight lines through one point.

The initial question, and the answer to it, can be generalized in various important directions. The most vigorous work on the subject has been done by Ragnar Frisch. Current debates on the foundations of the econometric methods and on the statistical time series have suggested to the present writer that complete clarity on the subject is certainly still far from universal and that a systematic and elementary treatment may be of some use if only to point out where the unsolved questions seem to lie.

A set of observed values of m variables $x_1, \cdots, x_m$ can be looked upon as a solution of m equations. The observed ith set, for example, may be the solution of a system of m equations

$$f_j(x_1, \cdots, x_m; a^j{}_{i1}, a^j{}_{i2}, \cdots, a^j{}_{is_j}) = 0, \quad (j = 1, \cdots, m) \tag{4}$$

where the a's designate parameters. Some of these equations may be postulated from theory while others express environmental conditions or (which is the same thing) causations unknown or irrelevant to theory. If the observed values vary from set to set, this must be due to changes in at least one parameter. Graphically, if none of the parameters changes from set to set, the "scatter" of, say, n observation points

$$(x_{i1}, x_{i2}, \cdots, x_{im}) \qquad (i = 1, \cdots, n)$$

reduces to one single point—the point of intersection of the m surfaces (4).

Suppose, then, that not all m equations (4) are "fixed." Then two questions arise. (1) Which are the equations (say, p of them in all) with the relatively stable parameters? (2) Supposing that these p equations have been identified, what are the conditions making possible the evaluation of their parameters?

The answer to the first question is sometimes provided from additional sources of information. For example, it may be taken that the equation connecting wheat consumption, wheat price, and national income is subject to negligible changes of the parameters involved, compared with the equation which expresses wheat crops as a function of price: because the weather which affects wheat crops fluctuates more strongly than the taste for this necessary and well-estab-

lished component of human diet. Such an assumption is additional to the information embodied in the statistics of prices, quantities, and incomes (used for the derivation of the equations themselves) and cannot be derived from those statistics.[9]

V. LOST DIMENSIONS

Unlike the first question (location of shifting parameters) the second question (possibility of measuring the stable parameters) is capable of an answer based on the nature of the theoretical system itself.

A simple case has been discussed at the beginning of the preceding section. As far as I know, more complicated cases have not been studied in similar detail, nor would such study present much interest here. The principles which we are going to discuss will be sufficiently elucidated by studying an extreme case, namely, by assuming that all parameters of the "stable" equations should remain strictly constant through the observations, while some or all parameters of any other equations change from observation to observation. Instead of estimating the parameters by the method of least squares or otherwise, we shall have to find them by exact solution. Graphically, we shall assume the points of the scatter to lie *exactly* on some of the theoretical surfaces; the surfaces will not have to be "fitted to" but to be drawn through the observed points. This simplifies considerably the presentation of what follows. It means, for example, that, instead of discussing the "normal equations" formed to obtain regression coefficients from the observations, we shall study directly the theoretical equations, with observed values substituted for the variables ("observation equations"). The somewhat conventional technicalities of the correlation analysis, irrelevant to our main purpose, are thus cleared out of the way.[10]

The problem treated at the beginning of Section IV, is thus reduced to this: of the two equations of the system, one—say the first one—is stable. Find its parameters. (Obviously, one observation is sufficient to find the one unknown parameter.) When extended to

[9] If measurable indicators could be introduced to represent "weather" and "tastes" (e.g., humidity and temperature for "weather" and some physiological characteristics of people for "tastes"), some idea as to the relative violence of the shifts in the respective parameters might be gained. Or, one can go further and use those indicators themselves as new variables of the system. In this augmented system, the demand and supply equations, each of them richer in variables than before, would be reckoned as "stable," while all shifts would be located in the equations which would describe the causation of humidity or of tastes and could be left out of the economic analysis.

[10] See, however, n. 11.

more than two variables, our problem assumes additional features. Consider first this example, not uncommon in economics:

$$
\begin{aligned}
z &= a\,x + b\,y + c \\
z &= g \log x + h \log y + k\,,
\end{aligned}
\tag{5}
$$

these being two "stable" equations in three variables; the third equation of the system has changing parameters and need not be written down explicitly. Regarding the parameters a, b, c, g, h, and k as unknown, we can find them from the six observation equations:

$$
\begin{aligned}
z_i &= a\,x_i + b\,y_i + c \\
z_i &= g \log x_i + h \log y_i + k\,.
\end{aligned}
\qquad (i = 1, 2, 3)
$$

We can, thus, draw a plane through the observations (x_i, y_i, z_i) and, independently, another plane through $(\log x_i, \log y_i, z_i)$.

As the next example, suppose that we have derived from theoretical considerations the system

$$
\begin{aligned}
z &= a \log x + b \log y + c \\
z &= g \log x + h \log y + k\,.
\end{aligned}
\tag{6}
$$

By substituting the observed values and solving with respect to the parameters, we draw again a plane through points $(z_i, \log x_i, \log y_i)$; but which of the two planes visualized by theory have we fitted, the one involving the coefficients a, b, c, or the one involving g, h, k? The question is similar to the one posed in an earlier part of the article, where the scatter was reduced to a point; here it is reduced to a line (straight line in the coordinates z, $\log x$, $\log y$). There is an infinite number of straight lines through a point and of planes through a straight line.

We have chosen on purpose an example where the original variables appear in equations in a nonlinear form; in this example, however, the case reduces itself, by a simple transformation of coordinates, to a linear one—to the problem of multicollinearity studied in detail in Frisch's *Statistical Confluence Analysis*. By extension, system (6) may be called "multiconform."

The case can, however, be stated more generally. Consider, for example, these two stable equations in three variables:

$$
\begin{aligned}
(x - g)^2 + (y - h)^2 + (z - k)^2 - r^2 &= 0 \\
x + a\,y + b\,z + c &= 0\,.
\end{aligned}
\tag{7}
$$

Geometrically, the scatter points lie on the intersection of a sphere

and a plane; that is, they lie on a circle. Now, given this circle, there can obviously be only one plane on which that circle can lie; but there is an infinite number of spheres passing through that circle (the Arctic Circle may be the Equator of some smaller Earth overlapping ours). The scatter is too poor in dimensions to enable constructing the sphere. Thus, geometrical intuition tells us that the coefficients g, h, k, and r cannot be found, while the coefficients a, b, and c can. As to the analytical test: by the conditions of our problem all the observation equations are compatible; it suffices, therefore, to pick out, in accordance with the number of the unknown parameters, four observation equations for the sphere:

$$(x_i - g)^2 + (y_i - h)^2 + (z_i - k)^2 - r^2 = 0 \quad (i = 1, \cdots, 4)$$

and three observation equations for the plane:

$$x_i + a\, y_i + b\, z_i + c = 0\,, \qquad (i = 1,\, 2,\, 3)$$

all other observations being redundant.

First consider the observation equations of the sphere. If the left-hand term of one of these equations—say, the first one—can be represented as the function of some of the remaining ones (i.e., if the left-hand terms are dependent functions), the first equation is redundant, and the system is indeterminate. To find whether the left-hand terms are, in fact, dependent, the following theorem can be used: the Jacobian determinant of m functions in m variables vanishes if and only if the functions are dependent. A Jacobian is a determinant each element of which is the partial derivative of one of the m functions, taken with respect to one of the m variables. On differentiating the observation equations of the sphere with respect to its unknown parameters, we have to see whether the determinant vanishes, i.e., whether

$$|-2(x_i - g),\, -2(y_i - h),\, -2(z_i - k),\, -2r| = 0\,. \qquad (i = 1, \cdots, 4)$$

Dividing each column by -2 and substituting (from the observation equations of the plane) $x_i = -(a\, y_i + b\, z_i + c)$, we obtain

$$|-a\, y_i - b\, z_i - c - g\,,\; y_i - h\,,\; z_i - k\,, r|\,. \qquad (i = 1, \cdots, 4)$$

The first column is reduced to a column of zeros if the other three columns are multiplied, respectively, by a, b, and $(a\,h + b\,k + c + g)/r$, and added to the first. Thus, if the observation equations of the plane are to be satisfied, the observation equations of the sphere form an indeterminate system, and the parameters g, h, k, and r cannot be found. But the parameters of the plane—a, b, c—can: the Jacobian

for the observation equations of the plane is (differentiating with respect to a, b, c):

$$| y_i \quad z_i \quad 1 |. \qquad (i = 1, 2, 3)$$

This does not vanish, even when the observation equations of the sphere are satisfied.

We can now give a more general formulation.

Suppose that, of the m *theoretical equations* (4) in m economic variables, p equations $(0 < p < m)$ are identified as having stable parameters only. Then n sets of observations yield each p sets of *observation equations,* each set consisting of n equations. The pn observation equations are:

$$f_{ji}(x_{i1}, x_{i2}, \cdots, x_{im}; a^j{}_1, \cdots, a^j{}_{s_j}) = 0, \qquad (i = 1, \cdots, n; j = 1, \cdots, p)$$

s_j being the number of parameters.

If $m = 2$, $p = 1$, and the only condition for the parameters to be obtainable from observations is $s_1 \leqq n$.

But if $m > 2$, $p \geqq 1$, and it becomes possible that whatever the number of observations as compared with the number of parameters, some of the sets of the observation equations become dependent and therefore indeterminate. This is shown by the vanishing of one or more of the Jacobians:

$$\left| \frac{\partial f_{ji}}{\partial a^j{}_k} \right|. \qquad (j = 1, \cdots, p; i = 1, \cdots, s_j; k = 1, \cdots, s_j)$$

Each such Jacobian refers to one of the p sets of observation equations but *must be evaluated by taking into account all the other sets.*

If $m = 2$, then $p = 1$, and the question of dependence of observation equations does not arise, although the question of identifying the stable equation of the system does. Therefore, when dealing with examples in two variables, the problem of variate parameters is easily understood, but the more general problem of "lost dimensions" keeps out of sight.

The results apply, of course, also to linear systems, or those transformable into linear ones—like in (6); the Jacobian of a set of linear observation equations in its reduced form (in $m + 1$ variables and parameters)

$$x_{i0} + b_1 x_{i1} + \cdots + b_m x_{im} + b_0 = 0 \quad (i = 1, \cdots, m + 1) \qquad \text{(B)}$$

is the determinant $|x_{i1}, \cdots, x_{im}\ 1|$. If another set of linear observation equations, in the same variables,

$$x_{i0} + a_1 x_{i1} + \cdots + a_m x_{im} + a_0 = 0 \quad (i = 1, \cdots, m+1) \qquad \text{(A)}$$

is to be satisfied together with the first, the above determinant vanishes, as is seen by multiplying its columns by $(b_1 - a_1)$, $(b_2 - a_2), \cdots$, and adding together, thus forming a new last column whose elements, as is seen by subtracting (A) from (B), are equal to zero. The vanishing determinant is, of course, common to both systems (A) and (B), and so the parameters of neither of them can be found.

Geometrically, the intersection of two planes is a straight line, but there is an infinite number of planes going through the same straight line.[11] Suppose, however, that one of the planes is tied to some condition; for example, it must be parallel to one of the coordinate planes or inclined to it at some given angle. There is only one such plane going through the line given by the equations. Analytically, this is the case if one of the coefficients in (B), say b_1, is known. While the determinant of the observation equations (A) has still the same form and vanishes for the same reason as before, the determinant of the observation system (B) becomes shorter by a row and column (because the term $b_1 x_{i1}$ is now a constant):

$$|x_{i2}, \cdots, x_{im} 1| \qquad (i = 1, \cdots, m)$$

and does not vanish.

Such cases frequently occur in practice, especially when the coefficient of one of the variables is known to be zero in one of the stable theoretical equations, so that the variable does not appear at all in that equation. Suppose, for example, the demand to depend linearly on price x and income y, and the supply to depend linearly on price x only, and both the demand and the supply equations to be "stable" compared with the missing third equation of the system. Equating demand = supply = z (say), we have (using again the reduced form)

$$z = a\,x + b\,y + c$$

$$z = g\,x + h\,.$$

The supply equation (a plane parallel to the y-axis) can be found, but not the other one.

When scrutinizing a system of economic equations, like Tinber-

[11] Instead of studying the interdependence between linear *observation equations* (B) as above, Ragnar Frisch studied the interdependence between the linear *normal equations* resulting from the assumption that the sum of squares of the left-hand terms of (B) is minimum. If the normal equations are dependent (collinear), the correlation determinant (i.e., the determinant whose elements are simple correlations between the variables) must vanish. In the "exact" case, discussed in the text, the sums of squares are zero and therefore minimum; and the correlation determinant vanishes. In the practice of "fitting" it never vanishes exactly.

gen's, it is not correct to discard it (as some critics do) on the sweeping accusation of the "interdependence of variables." As far as I see, the various theoretical systems used in his work stand the test of dimensionality. The same is true of at least some of the simpler attempts made by Schultz. He argues that the supply (z) of agricultural commodities probably responds not to current prices (x) but to prices of the preceding year (x'). Suppose both the demand equation $z = F(x)$, and the supply equation $z = f(x')$ are linear; no loss of dimensions need be feared, unless further relationships exist. In each of the two equations the coefficient of one of the variables, a different one in each case, has a known value, viz., zero, as in the preceding paragraph.

When the relationships involved are not linear in the variables but are linear in parameters and can therefore be made linear in variables by transformation, as in the cases (5) and (6) above, the same rule will apply: in equations where some parameters have a known value, others can be found, provided that the known parameters are not the coefficients of the same variable in any two equations. Also, if, in the example (7) (where the function involved is not a linear one), of the four parameters of the sphere one is given a priori, the other three can be found. This is easily shown by testing the Jacobians of the observation equations or by geometrical construction ("through a given circle draw a sphere of a given radius r" or "through a given circle draw a sphere the center of which has distance g from the yz-plane").

Thus, in the case of linear and certain nonlinear relationships, the addition of one further variable with an unknown parameter makes the equation where this addition is made indeterminable. If the additional variable is "time," we obtain the rule—as in Section IV which now appears as a special case—that the "unstable" equations are not determinable. Whether the rule can be extended to wider classes of equations[12] would be a problem of some practical importance, as it would make it unnecessary, in those cases, to evaluate the Jacobians of observation equations in order to find whether the parameters can be determined.

VI. CORRELATION BETWEEN AND WITHIN TIME SERIES

There is at present a strong reaction among economists against the overoptimistic application of statistical methods to economic time

[12] It has been pointed out to me by T. Haavelmo that such an extension is by no means obvious. See his paper presented at the Research Conference of the Cowles Commission, Colorado Springs, 1940.

series. The difficulties are, of course, due not to some mystical property of "time" but to the interdependence within the economic system. The dangers, not always clearly distinguished, are of two kinds: (1) dangers due to unsuspected additional relationships and (2) dangers due to relationships which, although assumed by the investigator, do not fit the framework of the sampling theory in its present form. The former type of dangers was described as "spurious correlation" and presented as such by Pearson and by Yule;[13] the latter was treated in Yule's article on "nonsense correlations."[14]

1. Suspicion of "spurious correlation" is expressed, for example, in Keynes's criticism of Tinbergen's method:

> Must we push our preliminary analysis to the point at which we are confident that the different factors are substantially independent of one another? . . . If we are using factors which are not wholly independent, we lay ourselves open to the extraordinary difficult and deceptive complications of "spurious" correlation . . . . . My mind goes back to the days when Mr. Yule sprang a mine under the contraptions of optimistic statisticians by the discovery of spurious correlation. In plain terms, it is evident that if what is really the same factor is appearing in several places under different disguises, a free choice of regression coefficients can lead to strange results.[15]

The last sentence suggests that Keynes has in mind the case discussed above (Sec. V) as the case of lost dimensions. Pearson's and Yule's emphasis was somewhat different. They chose as the simplest example the correlation between x and y when these variates have as a common factor a variate z. Obviously, a similar situation exists when z is not a common factor but a common additive or, still simpler, when $x = y + z$; this latter case has, in fact, been often used to explain the logical origin of the correlation coefficient,[16] because in this case the square of the correlation coefficient r^2 measures the "proportion of the variance of x explained by y."[17]

[13] K. Pearson, "On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs," *Proceedings of the Royal Society of London,* 1897; G. U. Yule, *Introduction to the Theory of Statistics* (1927), chap. xi.

[14] "Why Do We Sometimes Get Nonsense-Correlations between Time Series?" *Journal of the Royal Statistical Society,* 1926.

[15] *Economic Journal,* September, 1939, pp. 561–62.

[16] G. Darmois, *Statistique et applications,* p. 129; M. Ezekiel, *Methods of Correlation Analysis,* p. 375.

[17] A useful approximation formula for the general case is given by B. S. Yastremsky: If $x = x(z, u)$ and $y = y(z, v)$, where u and v are independent of each other and of z, then the correlation coefficient

$$r_{xy} = x_z y_z \sigma^2_z / \sqrt{(x^2_z \sigma^2_z + x^2_u \sigma^2_u)(y^2_z \sigma^2_z + y^2_v \sigma^2_v)},$$

where x_z, x_u, y_z, and y_v are partial derivatives, and the σ's denote the standard deviations. (Yastremsky and Chotimsky, *Theory of Mathematical Statistics* [in Russian] [Moscow, 1930], pp. 357–59).

The point about "spurious correlation" is not the *existence* of some common cause but its *unsuspected* existence. If, on the contrary, the existence of the common cause is the very hypothesis we are testing (or a part of it) then there is nothing spurious about the correlation. As already stated, the investigator takes into account all the factors which may affect each predictand, grouping them into a system of theoretically meaningful relations. The danger of "lost dimensions" would, but need not always, arise if there is more than one relation between the variables, but this is another question (Sec. V).

2. A different question is the difficulty to apply ordinary sampling analysis in its present form to sequences of not independent observations. This serial interdependence may, again, be the result of the very relations the economist studies as his hypotheses. For example, if the supply of a given year determines the current price (demand function) but is itself determined by the price of the preceding year (supply function), then there is a relationship holding for any two successive prices. We saw in Section III that relations of this kind (eq. [1]) may constitute the very hypotheses the economist wants to test. The parameters of such a theoretical or "true" relation are the parameters he has to estimate. The attention of economic statisticians has been too often absorbed by the question of "removing" these relations.

Suppose, then, the variable—for example, price—has, in any two successive years, the values x_t and x_{t+1} obeying the simple relation

$$x_{t+1} = \beta x_t + \varepsilon_t , \qquad (t = 1 , \cdots , n) \tag{8}$$

where the "error" ε_t is uncorrelated with x_t. The coefficient β is the unknown proportion which must, on theoretical grounds, exist between the expectation of the price in a given year and the actual price in the preceding year. The statistical problem is to estimate β from the observed series of prices in $n+1$ years.

We note that the regression coefficient of x_{t+1} on x_t,

$$b = \frac{\sum (x_t - \bar{x}) (\beta x_t - \beta \bar{x} + \varepsilon_t - \bar{\varepsilon})}{\sum (x_t - \bar{x})^2}$$

$$= \beta + \frac{\sum (x_t - \bar{x}) (\varepsilon_t - \bar{\varepsilon})}{\sum (x_t - \bar{x})^2} = \beta + b_{\varepsilon x} ,$$

does approach the true value as the number of observations increases and the value which the term $b_{\varepsilon x}$ on the right-hand side can exceed with any given probability, approaches zero (because of non correlation between ε and x): b is a "consistent" estimate of β. Further, the

sample regression coefficient b is distributed around β in the same way as the sample regression coefficient ($b_{\varepsilon x}$) of ε on x is distributed around zero. But this distribution is not the distribution proper to the sample regression coefficient in the classical case. The difference can be illustrated on the following simple model. Imagine an indefinitely large number of urns, each bearing on the outside one of the existing integer numbers, positive or negative, no number being repeated. Each urn contains an indefinite number of slips inscribed, for the urn r, with the numbers $\beta r+\varepsilon$, where ε is a normally distributed random number with zero mean, and a variance which is the same for all urns. If we choose n urns, record the value x_r of the slip drawn from any urn r, and calculate the regression coefficient between x_r and r for the sample of n drawings, and if, for the same n urns, such samples are repeated indefinitely, then the various regression coefficients will be distributed around β in the classical way. Suppose, however, that merely the first urn is the same in each sample, while the other urns are chosen in the following way: The number written on the slip drawn from the first urn (i.e., x_1) will be the number of the second urn, etc., each successive urn being determined by the number on the slip drawn from the preceding one. We have, then, our relationship (8): the urn number, which is the expectation of the slip number, is made to depend on the observed previous slip number. I do not know whether the distribution of sample regression coefficients in this and generalized cases has been studied; and whether other, more appropriate estimates have been considered. In the interest of economic work such studies would be highly desirable.[18]

If the errors in (8) were all zero, that equation would describe the "true path" (or "trend"?) of x through time. The path is exponential for $\beta > 0$, and periodical—with a period of 4 time units—if $\beta < 0$: as is familiar to the students of the "cobweb" in economics. Similarily, if instead of (8), we had

$$x_{t+1} = x_t + \alpha + \varepsilon_t \,, \qquad (t = 1, \cdots, n; \alpha \text{ constant})$$

the trend would be linear. A consistent estimate of the rate of increase α is $a = \frac{1}{n} \sum_{1}^{n} (x_{t+1} - x_t)$.

So long as the trend, or path, is the very relationship we are trying to estimate, there is nothing dangerous in the fact of the serial

[18] Cf. J. Neyman, *Lectures and Conferences on Mathematical Statistics* (Washington, 1938), pp. 109–24; M. M. Flood, "Recursive Methods and the Analysis of Time Series," Cowles Commisson for Research in Economics, *Report of Fourth Annual Research Conference* (Chicago, 1938).

interdependence, except the fact that the relevant statistical distributions await the mathematician who would calculate them.[19]

VII. SUMMARY

To test theories describing the interdependence of economic variables, empirical equations of the Schultz type are not a useful tool. On the other hand, their use for prediction and policy—assuming a reasonable degree of uniformity in the economic world—is not as endangered by the economic interdependence and the nature of time series as is sometimes felt. The problems of "variate parameters" and the more general one of "lost dimensions" are studied. The meaning of "spurious correlation" is discussed; the difficulties inherent in series of interdependent observations are shown to be due merely to the present lack of knowledge of the relevant statistical distributions.

[19] Of this nature is also the problem treated in Yule's article on "nonsense-correlations." The correlation coefficient between sin t and sin $(t + \text{lag})$ has a U-shaped distribution if the samples are taken over small parts of the cycle: the probability of a comparatively wrong estimate of the correlation coefficient is higher than the probability of a comparatively correct estimate. This shows, incidentally, the danger implied in the method of estimating the true time lag by choosing the one which makes the correlation coefficient a maximum. But better methods could be devised.

WEIGHTED REGRESSIONS IN THE ANALYSIS OF ECONOMIC SERIES

JOHN H. SMITH
University of Chicago and the Cowles Commission

IT IS sometimes appropriate to specify that a set of observations on k related variables constitutes a sample selected at random from a k-variate universe characterized by linear regressions. When the arrays of the dependent variable are normal and homoscedastic, the elementary regression coefficients are maximum likelihood estimates of the corresponding parameters of the single-valued linear relationship which exists among the independent variables and the array means of the dependent variable. A different type of regression is appropriate when the set of variables satisfies a single-valued relationship except for random normally distributed errors in all the variables. The regression coefficients are taken as estimates of parameters of such a single-valued relationship. Schultz called this type of weighted regression the mutual regression.[1] The set of techniques known as confluence analysis developed by Frisch[2] and the restatement and extension of Frisch's arguments in terms of sampling theory by Koopmans[3] are also related to this type of regression. For either type of specification the variables may be logarithms, or any known functions, of observed economic series.

This essay deals with a few of the most important arguments used to justify the use of weighted regressions with special reference to (1) simple specifications of universes and conditions of sampling for which weighted regressions are well defined and to (2) the nature of problems in economic research which may legitimately be identified with the determination of one or more parameters of a weighted regression.

THE NATURE OF ERRORS IN ECONOMIC VARIABLES

The use of weighted regressions has been advocated for macro-

[1] Henry Schultz, *The Theory and Measurement of Demand* (Chicago: University of Chicago Press, 1938), pp. 148–49.
[2] Ragnar Frisch, *Statistical Confluence Analysis by Means of Complete Regression Systems* (Oslo: Universitetets Økonomiske Institutt, 1934).
[3] T. Koopmans, *Linear Regression Analysis of Economic Time Series* (Haarlem: De Erven F. Bohn N.V., 1937).

dynamic studies in which the variables are index numbers of economic phenomena for a whole country. It is difficult to define, even in terms of complete and accurate data, a set of perfectly correlated index numbers for a country like the United States. Yet the use of weighted regressions seems to imply that deviations from such a set of ideal series are *the* "errors" in a set of observations on k related index numbers. Consistent use of this specification requires that time, when used as an index number to represent the effects of neglected variables, should be assumed to contain an important error component. However, time is usually specified to be without error.[4] Schultz discussed the advisability of specifying stochastic relationships like those between heights of fathers and sons, but he preferred the specification of closed sets, except for errors, because single-valued relationships are indispensable in economic theory.[5] The relevance of this reason is worthy of examination.

A logically complete model economy can be described in words or in terms of mathematical symbols. At best this model resembles the relevant characteristics of the real world as closely as is consistent with the feasibilty of obtaining a solution for an economic problem. The unconditional acceptance of a solution based on a model economy implies that abstractions which are necessary in order to describe a manageable model are unimportant. However, it is usually necessary to revise model solutions in order to allow for the effects of problem conditions which the model does not describe adequately.

The use of single-valued functions in an economic model is easy to justify on the basis of simplicity alone when theorems based on postulates of the model are revised before being applied to real problems. Improvement in the degree of correspondence to the institutional setting of the real world introduced by substituting stochastic for exact relationships is less important than sacrifice of simplicity—an indispensable characteristic of the model. On the other hand, in statistical studies the specification appropriate for the use of elementary regressions leads to simpler methods of analysis and interpretation.

Neither Frisch nor Koopmans discusses the relative usefulness in economic research of weighted regressions as compared with elementary regressions. Both writers imply that regression analysis of economic series is always used to estimate the parameters of single-valued relationships which would exist in the absence of errors. Contrary to opinions which might be formed on the basis of available

[4] Schultz, *op. cit.*, pp. 717–18.
[5] *Ibid.*, pp. 146–48.

treatments of weighted regressions, the presence of errors in all the variables is not sufficient in itself to establish the applicability of the weighted regression. If the purpose of the investigation is to estimate the value of the dependent variable from future observations on the independent variables, it can be shown that the elementary regression is superior to the weighted regression even when the weighted regression coefficients are maximum likelihood estimates. The best case for the use of the weighted regression can be made when independent variables are controllable in the sense that something can be done to make an independent variable larger or smaller than it otherwise would have been. A change in tariff rates is a possible example. In general, it would seem that the relevant relationship is that which expresses the variable in which changes are to be predicted as a function of the controllable or observable components of related series.

WEIGHTED REGRESSIONS FOR A SIMPLE SPECIFICATION

Let each of a set of observed series be a linear function of $r + 1$ normally distributed component variables of the form

$$x_i = a_{i1}y_1 + a_{i2}y_2 + \cdots + a_{ir}y_r + c_iv_i\,. \tag{1}$$

The coefficient of each of the y's differs from zero in at least two of the x's, while each of the v's is peculiar to only one x and is usually specified to be the error component or disturbance. In case it is more appropriate to specify that the error components in a set of variables are correlated, one or more terms in the y's can be added to the v's and the sums can be specified to be errors which are, of course, correlated.[6] This is a reasonable specification when two or more series have been deflated by the same index number, but, in most cases, it merely increases the difficulties which always arise in the process of defining errors. It is usually better to treat correlated errors as the effects of neglected variables in order to place emphasis on whether or not the effects of omitted variables are really unimportant.

When the relationship among a set of variables is not perfect, it is impossible to set limits which will include the value of the coefficient of regression of the dependent variable on any independent variable when an additional variable is added to the set. For convenience consider any regression solved for x_1 in a set of seven variables. Let x_8 be constructed in such a way that each observation is proportional to the distance, in a direction parallel to the x_1-axis, from the corresponding

[6] Koopmans, *op. cit.*, pp. 51–53.

observation point to any arbitrarily selected plane in seven-space. When the equation of the single-valued relationship in the new closed set of eight variables is solved for x_1 the coefficients are identical with the corresponding coefficients of the arbitrarily selected plane. Hence, any regression coefficient (elementary or weighted) may either increase or decrease without bound upon the addition of a neglected variable to the set. In view of this conclusion, the mathematical treatment of the specification of single-valued relationships can be limited to cases in which the k variables form a closed set. This is accomplished by specifying that the rank of the determinant of the coefficients of the y's in the system of equations (1) is r and that r is less than k. Geometrically this means that the single-valued relationship among the systematic parts is the intersection of $k - r$ planes in k-space. It is also convenient to specify that all variables are measured in units proportional to their *known* disturbance standard deviations. The specification of normality permits the use of deviations from sample means in terms of which the maximum likelihood solutions are most easily expressed. The justification for this choice of origin is the same as for the determination of elementary regression coefficients.[7]

THE METHOD OF MAXIMUM LIKELIHOOD AND SINGLE-VALUED RELATIONSHIPS

Maximum likelihood estimates are values which maximize the appropriate frequency density function at the sample point. For example, in random samples of n observations from a normal universe, deviations from the mean of the universe are considered to be errors of sampling. The frequency density is a decreasing exponential function of the sum of squares of sampling errors. When this function is maximized by minimizing the sum of squares with respect to the unknown mean of the universe, the arithmetic mean of the sample is obtained as the maximum likelihood estimate. Geometrically this is equivalent to minimizing a distance in n dimensions. Consider the sample point in relation to a similar point in which there are no errors. The coordinates of the latter point are all equal. Hence, the problem is to choose the point which is closest to the sample point subject to the restriction that all of its coordinates are equal. The locus of such points is a line whose direction cosines are all equal to the reciprocal of the square root of n because the sum of squares of any set of direction cosines must be unity.

The geometrical interpretation of the problem of estimating the

[7] Schultz, *op. cit.*, pp. 734–35.

mean of a normal universe is analogous to the problem of estimating the parameters of a single-valued relationship among a set of variables when the systematic parts are all proportional and all variables contain independent normally distributed random error components—the simplest case of multicollinearity. Since variables can always be expressed in units proportional to the known standard deviations of their disturbances, only the case of equal disturbance standard deviations need be discussed. In practice, the unknown disturbance standard deviations must be estimated. Consider a sample of n observations on k variables represented as a scatter of points in k-space. The single-valued relationship is a line in k-space whose parameters are to be estimated by the method of maximum likelihood. As in the case of the mean, maximum likelihood estimates of the coordinates of an errorless point determine the foot of the perpendicular from the sample point to the line of relationship. Since the parameters of the line are unknown, the distance from the sample point to the line must be expressed in terms of the unknown parameters of the line. In order to maximize the frequency density function for the sample as a whole, it is necessary to minimize the sum of squares of these distances with respect to the unknown parameters.

When the single-valued relationship is a plane instead of a line, similar arguments are applicable. When units are chosen proportional to the standard deviations of independent normally distributed error components, the problem of obtaining maximum likelihood estimates of the parameters of the relationship among the systematic parts is reduced to the minimization of the sum of squares of perpendicular deviations from the plane of relationship with respect to the unknown parameters. Geometrical representations of the solutions for such problems are lines and planes which depend only on the configuration of the observation points and are therefore independent of the choice of coordinate axes.

Let M represent the square array of variances and covariances among the x_i's which is known as the moment matrix. The characteristic equation is derived by subtracting m, an undetermined constant, from each of the diagonal entries of M and setting the determinant of the resulting matrix equal to zero.[8] Let M_j represent the characteristic moment matrix when the jth root in order of size is substituted for m. The general solution of the equations required for the minimization process for the case of proportional systematic parts shows that

[8] L. L. Thurstone, *The Vectors of Mind* (Chicago: University of Chicago Press, 1935), pp. 1–43. This is the simplest available treatment of mathematical topics related to the characteristic equation.

maximum likelihood estimates of the direction cosines of the line of relationship are obtained by normalizing the elements of any row or column of the adjoint of M_k. Similarly, normalizing the elements of any row or column of the adjoint of M_1 determines a set of maximum likelihood estimates of the parameters of the plane of relationship in normal form when the systematic parts form a closed set with no closed subsets.[9] These solutions are determinate only when the appropriate characteristic roots are distinct.

When all characteristic roots are distinct, normalization of any row of the adjoint of every M_j, where j ranges from 1 to k, determines the coefficients of k mutually perpendicular planes through the origin. The same sets of coefficients are direction cosines of normals to the corresponding planes at the origin. Let the line based on the ith characteristic root be taken as the axis of a new variable z_i. The z's are uncorrelated linear functions of the x's—perpendicular deviations from the k planes—in terms of which it is possible to summarize briefly the properties of the lines and planes of closest fit. The variance of normalized linear functions of the x's is a minimum for z_1 and a maximum for z_k. The variance of z_i is a minimum subject to the condition that all planes, $z_j = 0$, are perpendicular where j ranges from 1 to i. In the case of multiple characteristic roots, some of the unknowns can be chosen arbitrarily in the general equations of minimization, but this is seldom necessary in practice.

When it is appropriate to specify that the systematic parts of a set of variables form a closed set with no closed subsets, the maximum likelihood estimate of the single-valued relationship is the plane $z_1 = 0$. Advocates of weighted regressions take for granted that this type of regression should be used except in cases of multicollinearity. The most extreme case of multicollinearity has been discussed. In intermediate cases the single-valued relationship among the systematic parts is the intersection of g planes in k-space where g is greater than one and less than $k - 1$. This relationship, determined by the method of maximum likelihood as before, is the intersection of the g planes $z_i = 0$ where i ranges from 1 to g. The coefficients of the functions which express the z's in terms of the x's are separately unreliable even under conditions favorable to the reliable determination of the intersection. The unreliability of the individual coefficients has led to the common statement that regression coefficients have no meaning in cases of multicollinearity where variables in the set satisfy more than one linear relationship. A common procedure in economic studies is

[9] Koopmans, *op. cit.*, pp. 47–53, 63–65.

to choose from the multicollinear set of variables that subset which most nearly approaches a closed set with no closed subsets. It is interesting to contrast this method with that of multiple-factor analysis in which multicollinearity is a necessary condition for success.[10]

When it is appropriate to specify that the systematic parts are proportional to each other, z_k is the best estimate of their common component according to the criterion of maximum likelihood. This is probably the most important case of the z's because the existence of multicollinearity among economic variables is common. This characteristic is an important factor in the explanation of the common practice of combining highly correlated variables into averages or index numbers, even when the differential behavior of some of the variables in the set is important for other purposes. This also serves to reduce the number of variables, but in many cases unreliability of regression coefficients is sufficient in itself to justify the use of index numbers. Estimates of the separate effects of independent variables in a highly correlated subset are extremely unreliable even when the combined effects of the whole group of variables is great. In such cases it may be advisable to abandon the attempt to estimate the separate effects of variables in nearly closed subsets and to specify that the systematic parts of variables in the subset are proportional where each series contains a normally distributed disturbance. The function z_K can then be used as the most reliable linear function to represent the economic variable imperfectly measured by every member of the multicollinear set. The disturbance standard deviations which are assigned for the purpose of computing z_k should be determined primarily by the degree to which each variable seems to represent the desired economic variable.

The original specification describes the x's as linear functions of r component variables which are common to two or more x's where r is the rank of the moment matrix of the systematic parts and the single-valued relationship among the systematic parts is the intersection of $k-r$ planes in k-space. The observations on the systematic parts are restricted to a region of r dimensions defined by this intersection. For example, when r is unity, the observations on the systematic parts lie on a line defined in terms of maximum likelihood estimates as the axis of z_k. Since z_k is influenced only by variation parallel to this line, it is superior to any other linear function of the x's for the purpose of representing the single common factor. When r is two, observations on the systematic parts are restricted to a region of two dimensions. The variables z_k and z_{k-1} represent the two com-

[10] Thurstone, *op. cit.*, pp. 72–77.

mon factors in a certain sense. They are maximum likelihood estimates of uncorrelated functions of the x's influenced only by variation parallel to the region of two dimensions which contains all the observations on the systematic parts. Neither z_k nor z_{k-1} has any simple interpretation with respect to either of the two common factors. When both z_k and z_{k-1} are used as independent variables in regression analysis, their combined effects differ from the combined effects of the common factors chiefly because of errors in estimating parameters by the method of maximum likelihood. Devices used in multiple-factor analysis for the estimation of the various common factors in terms of the x's are not of interest for present purposes. The foregoing argument which can be extended to apply to larger values of r for any problem in which such extension is appropriate is consistent with the other aims of the multiple-factor method.

NUMERICAL APPROXIMATION OF PARAMETERS OF SINGLE-VALUED RELATIONSHIPS

When a specified single-valued relationship is a straight line, maximum likelihood estimates of its direction cosines are proportional to the elements of a row of the adjoint of M_k. When the number of variables is large, the labor of computing these estimates by ordinary processes is considerable. It is necessary to compute all principal minors of all orders to obtain the characteristic equation. The numerical value of the largest characteristic root must then be approximated by Horner's or some similar method. Finally, this root is substituted in the characteristic matrix and the elements of a row of the adjoint are computed and normalized. This direct method can be replaced by a simpler iterative process which is especially useful when the number of variables is large. Except for choice of units, the iterative method is equivalent to Hotelling's method of principal components.[11] When the rows of the singular matrix M_k are multiplied by numbers proportional to the corresponding elements of any row or column of its adjoint, the column sums vanish. When the moment matrix M is multiplied by the same vector, the column sums constitute a vector which is m_k times the original vector. This follows from the fact that diagonal entries of the two matrices differ by m_k. When the original vector is only approximately proportional to a row of the adjoint of M_k, the process is iterative. When successive vectors are normalized, their elements converge toward maximum likelihood estimates of di-

[11] Harold Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, XXIV (1933), 417–41, 498–520.

rection cosines of the line of relationship which corresponds to the axis of Hotelling's first principal component. The rapidity of convergence of the iterative process is increased by squaring the moment matrix several times before multiplying by the first vector.[12] The result obtained from 2^t iterations applied to the moment matrix can be secured by a single iteration applied to the matrix obtained by t successive squarings.

Hotelling's iterative processes are applicable to the estimation of parameters of the line of relationship. They can be adapted for use in estimating the parameters of the weighted regression—a plane of relationship. When a number larger than m_k is subtracted from each of the diagonal entries of the moment matrix, the characteristic roots of the resulting matrix are all negative. Corresponding characteristic roots of the two matrices differ by a constant amount, but their ranking in order of absolute size is reversed. Since the iterative process converges toward a set of direction cosines based on the largest root in absolute size, coefficients of the weighted regression are obtained.

The original specification of the nature of the x's determines a normal k-variate universe. When the rank of the moment matrix of systematic parts is less than $k-1$, the set of systematic parts is multicollinear. Any such universe having nonvanishing disturbances can be transformed into a set of variables whose systematic parts form a closed set with no closed subsets by changing the specification so that all v's except v_1 are specified to be systematic parts. This implies that the systematic part of x_1 is the first elementary regression function. Similarly, any normal universe can be expressed in terms of the extreme case of multicollinearity, but it is usually necessary to specify correlated disturbances. Unnecessary use of correlated errors is avoided by expressing the x's in terms of the z's and specifying that terms in z_k are systematic parts.

The foregoing examples show that the type of relationship among the systematic parts depends on the choice of components which are specified to be errors. This principle has important practical applications. When the purpose of a study is to derive an estimating equation, it is irrelevant whether or not the available series have systematic parts which are perfectly correlated. The estimating equation should be derived in a manner consistent with its intended use; the methods of estimating its parameters should be the same as would be appropriate if there were no errors in the independent variables. An observed value of the dependent variable should be considered as a

[12] H. Hotelling, "Simplified Calculation of Principal Components," *Psychometrika,* I, 27–35.

random sample from an array of values associated with a given set of observations on the independent variables. The research worker should specify that the mean of any array of values of the dependent variable is a single-valued linear function of the independent variables. When such arrays are normal, maximum likelihood estimates of the required coefficients are the elementary regression coefficients.

Maximum likelihood estimates of any set of mathematically independent parameters of a normal k-variate universe can be obtained from a random sample of n observations where $n-k$, the number of degrees of freedom among the residuals, must be greater than zero—preferably considerably greater. One complete set of parameters consists of the means, the standard error of estimate, the regression coefficients (excluding the constant term), and the sampling variances and covariances of the regression coefficients. These sampling variances and covariances are parameters of the normal k-variate simultaneous distribution of the elementary regression coefficients where successive samples contain the same values of the independent variables.[13] The sampling distribution of a particular regression coefficient when sampling is unrestricted is the sum of the separate normal sampling distributions derived on the basis of particular sets of values for the independent variables. This distribution is leptokurtic and is inconvenient for practical purposes.[14] The conditions of stratified sampling simplifies the distribution and, at the same time, permits a simple derivation of an exact test of significance for the more general case.

When successive samples are selected from the same arrays, the sampling distribution of any regression coefficient in units of its estimated standard error is the t-distribution where the number of degrees of freedom is $n-k$. This statement which is relatively easy to establish for any single set of values of the independent variables[15] is also true for all possible sets when (1) the universe is continuous, (2) the arrays of the dependent variable are normal and homoscedastic, and (3) the regression is linear. This implies that no subset of the independent variables is closed by definition because the elementary regression in the universe is indeterminate in that case. For all determinate cases, the preceding statements follow from the fact that no closed subsets occur in sample values of the independent variables unless all samples contain at least one closed set. Absence of

[13] Schultz, *op. cit.*, pp. 720–26.

[14] Karl Pearson (ed.), *Tables for Statisticians and Biometricians* (London: Biometrika Office, University College, 1934), Part II, pp. cxliv–cliv.

[15] John H. Smith, *Tests of Significance: What They Mean and How To Use Them* (Chicago: University of Chicago Press, 1939), pp. 77–86.

closed sets depends on the property of continuity. For any sequence of finite universes which approaches a normal k-variate distribution, the limit of the relative frequency of samples containing closed sets among the independent variables is zero. Hence, the t-test is an exact test of significance on the basis of a specification which determines a nondegenerate normal k-variate universe.

CONFLUENCE ANALYSIS

The set of techniques known as confluence analysis was devised by Frisch for use in regression analysis.[16] Its most valuable device is the bunch map—one graph for every pair of variables in every subset. For example, a graph representing relationships between x_1 and x_2 in any subset of g variables is a set of lines radiating out from the origin in the first and fourth quadrants. The slopes of these lines are the coefficients of x_2 where the g elementary regression equations are solved for x_1 and all variables are expressed in standard-deviation units. The number i is placed near the end of the line or beam which is based on the ith elementary regression. The coordinates of the end of the ith beam are the ith elements of the second and first columns of the adjoint of the correlation matrix with appropriate signs.

Multicollinearity among a set of variables contributes toward shortness and wide spread among sets of beams in a bunch map. Hence, two criteria of desirability in bunch analysis are tightness of the bunch and absence of extremely short beams. Since the coordinates of the ends of beams of any bunch are proportional to the corresponding partial correlation coefficients, the necessary and sufficient conditions for a given degree of tightness and absence of short beams is that the absolute values of the highest order partial correlation coefficients must be sufficiently high. Since this is exactly the same condition which leads to high values of "Student's" t—a single-valued function of a partial correlation coefficient of highest order—the criteria of tightness of the bunch and absence of short beams correspond more closely to the common tests of significance than one might suppose after reading:

> In other words, they [tests practically equivalent to t-tests] will be entirely meaningless as tests of the "significance" of the regression coefficients. If we nevertheless use them we actually take a number drawn out of one hat as an expression for the "significance" of some other number drawn out of another hat.[17]

Although confluence analysis has much in common with the use

[16] *Op. cit.*
[17] *Ibid.*, p. 190.

of "Student's" t, bunch analysis has several advantages for the purpose of detecting characteristics which are commonly associated with samples from a multicollinear universe. The full bunch map is a graphic presentation of relationships among all possible subsets of variables. This facilitates the choice of a subset on the basis of a given set of criteria. The t-test is usually applied to only those regression coefficients of highest order which involve the dependent variable and the levels of significance used are often too inflexible for the purposes of the problem which Frisch had in mind. On the other hand, the tilling tables on which complete bunch maps are based involve so much additional labor that many who use bunch analysis for the complete set of k variables omit maps for all subsets. In this case, there is an actual saving in labor spent in computation as compared with the use of the t-test. Bunch maps for use in preliminary investigations are easily made. They probably require less effort than is necessary for computation of test criteria, but bunch maps are more expensive to publish. However, the bunch map presents more extensive information than is contained in the results of the usual tests of significance. Of course, no probability integrals are provided but probability integrals for simple t-tests are single-valued functions of the appropriate slopes and some idea of their magnitude can be gained from the appearance of the bunches.

The most serious disadvantage of bunch analysis is that mental allowance must be made for the number of observations in the sample. Too much reliance is likely to be placed on results obtained from small samples. Waugh's studies of price flexibility reported by Frisch are good examples.[18] In these two studies annual observations covering a period of thirteen years were used. Seven variables are included in each study, of which two are first differences and one is time as a catch-all variable. Hence, there are only twelve observations on seven variables and only five degrees of freedom among the residuals. In regression analysis of serially correlated economic time series this is a small sample. Even for random samples from a normal k-variate universe twelve observations on seven variables are hardly enough to make reliable tests possible. It is unfortunate that Frisch made no reference to the extremely small number of observations on which his extensive bunch analyses are based. However, the studies serve a useful purpose in illustrating an important principle of sampling which is likely to be neglected in bunch analysis.

When the specification states that samples are selected at random

[18] *Ibid.*, pp. 147–83.

from a normal k-variate universe, alternative hypotheses differ only in the magnitudes assigned to one or more of the $k(k+3)/2$ parameters of the universe. The common tests of significance are highly useful when a choice is to be made among such alternative hypotheses, provided that they differ in such a way that the test to be applied is the most powerful test.[19] For example, the t-test is the most powerful test which can be applied to a given regression coefficient when the values of that coefficient specified by alternative hypotheses are all greater or all less than the sample coefficient tested and all other parameters are the same in all alternative hypotheses.

Tests comparable in scope and efficiency with the common tests of significance under favorable conditions are not available for "realistic" specifications for regression analysis of economic time series. Alternative hypotheses for regression analysis of time series should include provision for neglected and potentially important variables as well as the processes which lead to nonsense relationships so often encountered in regression analysis of economic time series. These differences are not alone in magnitudes but also in the structural relationships of the universe which are absent from the simple specifications on which tests of significance or measures of reliability are based. These important characteristics of alternative hypotheses are "specified" out of existence by Koopmans when he states that "a repeated sample consists of a set of values which the variables would have assumed if in these years the systematic components had been the same and the erratic components had been other independent random drawings from the distribution they are supposed to have."[20] It is true that the t-test can be modified for use when the systematic parts are constant, but the modification of the specification does not add to the effectiveness of the test when a choice is to be made among alternative hypotheses which are plausible from an economic point of view—the only useful purpose of any test of significance.

In view of the fact that the major part of the variation of a time series is often identified statistically as the systematic part, the consequence of specifying that such components may be serially correlated is of some importance. Specifications which place no restriction of randomness on the systematic parts but offer no solution for problems related to dynamic changes are only superficially realistic. No statistical specification should be chosen in preference to a simpler

[19] Jerzy Neyman, *Lectures and Conferences on Mathematical Statistics* (Washington: Graduate School of U.S. Department of Agriculture, 1938), pp. 33–48.

[20] *Op. cit.*, p. 7.

one on the basis of resemblance to problem conditions for which neither offers any useful solution. Simplicity of specification is important when the same modification of results must be made in any case.

The work of Frisch and Koopmans has called attention to unrealistic specifications on the basis of which misleading interpretations of the common tests of significance are often made. On the other hand, the devices suggested by these writers are sometimes applied when simpler methods are superior. Since no set of techniques is universally applicable, methods of statistical analysis should be chosen on the basis of important problem conditions.

THE DISTRIBUTION OF HIGHER-ORDER INTERACTIONS

HARRY PELLE HARTKEMEIER
University of Missouri

THE analysis of variance was developed primarily to aid in the analysis of experimental data—that is, data collected by or under the direction of the statistician. The experiment was usually designed carefully, and considerable control was exercised over the conditions under which the data were obtained. When one attempts to apply the analysis of variance to data in the field of economics and business where experimentation and control are usually very limited, several difficulties are encountered. Data available in published form seldom provide the statistician with anything similar to replication in agriculture or biology, and consequently higher-order interactions must usually be resorted to as a measure of random variation. This is the best measure available for what the biologists call experimental error.

For example, the analysis of variance might be applied in a study of the money market. The existence of a national money market, in spite of the variation in the rate of interest charged customers in different localities, has been discussed by economists and bankers.

Such evidence as is available, therefore, indicates that fluctuations in rates charged customers are not governed directly by changes in the local discount rate which member banks must pay for Federal reserve bank accommodation nor by changes in the amount of local accommodation of this character which they find it necessary to obtain by borrowing. In addition to the fact, therefore, that member banks give precedence to customers' loans as compared with open market loans, they also seem inclined to charge customers the same rate without direct regard to whether their individual indebtedness at the reserve banks is increased thereby, or whether the rate which they have to pay on this indebtedness is changed. In contrast to these generalizations based upon local evidence, rates charged customers in leading cities appear to fluctuate on a national scale with fluctuations in rates in the open money markets, and to reflect, in the same manner, changes in the aggregate national volume of borrowing at the reserve banks and, possibly, changes in discount rates in other districts. If this similarity of movement is not purely fortuitous it means, first, that fluctuations in rates charged customers originate mainly in national changes in credit conditions and, second, that they are, consequently, competitive in origin. The latter conclusion, if true, has far-reaching implications, for it means that rates charged customers, in spite of the differentials that exist between them, the secrecy and confidence

with which they are negotiated, and the apparent lack of formal interrelationship between borrowers and lenders in different parts of the country, are made, nevertheless, in a national money market in which the play of competitive forces finds considerable scope and expresses itself in fluctuations of rates on a national scale.[1]

The apparent similarity of movements noticed by W. W. Riefler after a study of data presented graphically was not established by quantitative analysis. It seems evident that, if there is a national money market, the fluctuations of the rates charged customers in different localities would have to reveal similar seasonal variation (if there is any seasonal variation in the rates charged customers), and the business cycle must have a similar effect upon the rates charged customers in different sections of the country. Whether or not such is the case could be established easily and quickly by applying the analysis of variance, if adequate data were available on the rates charged customers in different sections of the United States. The *Annual Report of the Federal Reserve Board for 1931* contains (p. 82) data on the rates charged customers in principal cities for the period 1919–31, inclusive. The data cover only three sections of the United States—New York City, eight other northern and eastern cities, and twenty-seven southern and western cities. The limitations of the data and the method of weighting to obtain averages are discussed at length by Riefler in *Money Rates and Money Markets in the United States.* In spite of these limitations, the analysis of variance with three criteria of classification was applied to the data by the present writer to see if any quantitative evidence of the existence of a national money market could be obtained. The results appear in Table 1.

TABLE 1

RATES CHARGED CUSTOMERS IN PRINCIPAL CITIES, 1919–31
(Units of 1 Per Cent)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Geographical sections	2	52.9177	26.4588
Months	11	.2733	.0248
Years	12	158.5046	13.2087
Interactions:			
Sections and months	22	.1704	.0077
Sections and years	24	7.9227	.3301
Months and years	132	21.5990	.1636
Remainder	264	3.4064	.0129
Total	467	244.7941	....

Five of the first six mean squares are significantly larger than

[1] W. W. Riefler, *Money Rates and Money Markets in the United States,* pp. 71–72.

the remainder, or triple interaction, which is the best available measure of random fluctuation.[2] The mean square 26.4588 indicates that the rate charged customers is not the same in all sections of the country covered by the data. However, this fact does not prove that there is no national money market, for, as Riefler has pointed out:

> If the interpretation is correct that competitive forces are responsible for fluctuations in customers' rates, it is evidently not a competition that completely destroys local differences in the supply and demand for funds. Despite the fact that fluctuations in customers' rates appear to be national in scope and competitive in nature, there are well-defined differentials of long standing between the rates which customers in different localities pay for their loans.[3]

If it can be proved that the rates charged customers in different sections fluctuate in a similar fashion about their respective levels, it may still be possible to prove that there is in reality a national money market.

The mean square .0248 indicates that there is a seasonal variation in the rate charged customers. It is true that it is not a very large seasonal variation, but the mean square for months is significantly larger than that for the remainder, which is a measure of unexplained and/or random fluctuation. This does not indicate that the rate charged customers changes every month, but it does indicate that in at least one, and probably several, months the rate charged customers is different from the rate in other months.

The mean square for years, 13.2087, indicates merely what everyone knows—that the rate charged customers is not the same every year. This mean square really measures the combined effect of both secular trend and the business cycle, but in this case there is not a pronounced secular trend,[4] so the mean square for years measures principally the effect of the business cycle. These three mean squares are interesting, but they have been discussed principally because the analysis of variance is a fairly new technique. The interactions have a more important bearing upon the subject of this paper.

The mean square for the interaction between two of the criteria of classification—sections and months, .0077—is not significantly larger or smaller than the mean square for the remainder. This indicates that the seasonal variation is not significantly different in the

[2] For a test of significance see R. A. Fisher, *Statistical Methods for Research Workers* (2d ed.; London, 1928), pp. 194–96, and G. W. Snedecor, *Statistical Methods* (Ames, Iowa, 1938). For a discussion of the use of triple interaction as a measure of random fluctuation see R. A. Fisher, *The Design of Experiments* (London, 1935), pp. 115–17.

[3] *Op. cit.*, pp. 72–73.

[4] There is a slight negative, or downward, secular trend over the entire period covered.

different sections of the country studied. Stated in another way, the seasonal variation in the rates charged customers in different sections of the country varies from section to section only so much as would be expected on the basis of the random variation present in the data. This fact provides some evidence of the existence of a national money market, for within the year the rates charged customers in different localities vary about their own (different) levels in a similar fashion. This situation furnishes some evidence regarding the mobility of funds seeking investment. While it is evident that funds are not mobile enough to wipe out entirely the differences in rates charged, these differences due to local conditions are the result of a careful evaluation of factors peculiar to each locality, and competition holds these differences constant throughout the seasonal swing. This is even more remarkable in view of the fact that the seasonal variation itself is not constant throughout the period covered by the data. The change in the seasonal variation is indicated by the mean square for the interaction between months and years, .1636, which is significantly larger than the remainder. This mean square measures the extent to which the differences among the monthly rates change in magnitude from year to year, and in this case these differences among the monthly rates change more than would be expected on the basis of the random variation present in the data.

Only one mean square in Table 1 has not been discussed. It is .3301, and it measures the interaction between sections and years. Since this mean square is significantly larger than the remainder, it is evident that the business cycle does not have the same effect upon the rates charged customers in different sections of the United States. This mean square provides some evidence against the existence of a national money market from a long-run point of view. This mean square is reasonable, for depressions force down the rates charged customers in the North and East more than in the South and West. The amplitude of the business-cycle fluctuations is larger in the data for New York City than it is in the data for southern and western cities.

The conclusion seems to be that there is a national money market from a short-time point of view (for periods of one year or less), but from a long-run point of view (for periods longer than one year) the force of competition is not strong enough to cause similar business-cycle fluctuations in the rate of interest charged customers in different localities.

It may occur to the reader that the business-cycle fluctuations in different sections of the United States may be similar if relative fluc-

tuations rather than absolute fluctuations are considered. In order to investigate this possibility, the analysis of variance was applied to the logarithms of the data for the rates charged customers. The fluctuations in the logarithms are proportional to the relative fluctuations in the original data. Table 2 presents the results of this analysis.

TABLE 2

RATES CHARGED CUSTOMERS IN PRINCIPAL CITIES, 1919–31
(Logarithms of Units of 1 Per Cent)

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Geographical sections	2	.357773	.178886
Months	11	.002427	.000221
Years	12	.944203	.078684
Interactions:			
Sections and months	22	.001341	.000061
Sections and years	24	.067807	.002825
Months and years	132	.137849	.001044
Remainder	264	.026967	.000102
Total	467	1.538367	....

An examination of this table reveals that five of the six mean squares are significantly larger than the remainder. These are the same five mean squares that were found to be significant in Table 1. The principal effect of using logarithms is a reduction in the mean squares for years and sections which indicates that the relative fluctuations from year to year and from section to section are smaller than the absolute fluctuations.

The conclusions that may be reached by such analyses are largely dependent upon whether the remainder is a good measure of random variation. If the remainder is a good measure of purely random variation, it is necessary for triple interaction to cause fluctuations which are normally distributed. This possibility may be studied in detail by computing a table of values (similar to the original data) in such a way that these hypothetical data will reveal the same variation as the original data so far as the first six mean squares are concerned but contain no remainder or unexplained variation. Such a computation may be accomplished in a manner similar to the procedure used when one item in a table is missing.[5] Since the only difference between the original data and the hypothetical data is the presence of triple interaction in the original data, the hypothetical data may be subtracted from the original data in order to obtain a set of observations which have the same variance as the remainder and at

[5] See G. W. Snedecor, *Analysis of Variance* (Ames, Iowa, 1934), p. 40; and C. H. Goulden, *Methods of Statistical Analysis* (New York, 1939), pp. 263–65.

the same time contain no variance due to the principal components recognized by the three criteria of classification or the two way interactions between them. Table 3 presents a frequency distribution of the differences between the original data and the hypothetical data. Logarithms were used because it seems more appropriate to consider relative fluctuations rather than absolute fluctuations.

TABLE 3

TRIPLE INTERACTION OF RATES CHARGED CUSTOMERS IN PRINCIPAL CITIES, 1919–31 (Logarithms of Units of 1 Per Cent)

CLASS INTERVAL	FREQUENCY			
	Observed		Normal	
−.33 —* −.27	1	3	.07	1.03
−.27 — −.21	2		.96	
−.21 — −.15	14		7.98	
−.15 — −.09	30		37.11	
−.09 — −.03	98		98.29	
−.03 — +.03	175		141.21	
+.03 — +.09	97		114.74	
+.09 — +.15	38		52.27	
+.15 — +.21	9		13.39	
+.21 — +.27	3	4	1.85	2.00
+.27 — +.33	1		.15	
Total	468		468.02	

* Means "up to but not including."

When the chi-square test of the goodness of fit was applied to the data in Table 3, the value of chi-square was found to be 27.95. With six degrees of freedom this value of chi-square would arise due to chance alone less than once in a hundred random samples if the distribution under consideration were really normal, so it is best to conclude that the distribution is not normal. The triple interaction contains too many large deviations—the frequencies in both tail ends of the distribution are too large—in view of the general shape of the curve. These large deviations force the normal distribution fitted to the observed frequencies to have a larger variance than it should have in order to fit the center of the distribution properly and make the observed distribution look leptokurtic. This situation is confirmed by the value of β_2 which is 4.29. However, the distribution is practically symmetrical, for $\chi = -.0067$.

In this study triple interaction is slightly larger than it should be in order to provide a good measure of random fluctuations. This situation does not throw any doubt upon the conclusions reached above

by the analysis of variance when significant variation was found to be present, for with a smaller measure of experimental error (or random fluctuations) the significant mean squares would be still more highly significant or reliable. It is only the mean square which seemed to be nonsignificant which might mislead a research worker, for with a smaller mean square for the remainder (or triple interaction) the F ratio for the mean square .0077 (Table 1) or .000061 (Table 2) might prove to be significant.

IV

ECONOMETRICS

THE EMPIRICAL DERIVATION OF INDIFFERENCE FUNCTIONS

W. ALLEN WALLIS AND MILTON FRIEDMAN[1]
Stanford University and University of Wisconsin
National Bureau of Economic Research

ALTHOUGH the indifference function has become the keystone of the theory of consumer choice, empirical workers in the field, even those most thoroughly familiar with the niceties of indifference analysis, have ignored it or, in a few instances, dragged it in as a more or less irrelevant afterthought. Thus Professor Schultz, despite his brilliant presentation of indifference analysis in *The Theory and Measurement of Demand,* finds almost no occasion to employ it in the statistical portions of the book. Similarly, Allen and Bowley include in their monograph on *Family Expenditure* a discussion of indifference functions which has only the vaguest relation to their statistical material.

A critical analysis of the problem of quantifying indifference functions reveals weaknesses and inadequacies in the theory which justify the aloofness of empirical workers. The present essay, after summarizing the role of the indifference function in economic analysis, considers both experimental and statistical techniques for deriving indifference functions empirically. The difficulties and contradictions encountered are then shown to be inherent in the logical structure of the theory. It appears, therefore, that for empirical investigations of consumer expenditures an alternative theoretical framework is required.

I. THE ROLE OF THE INDIFFERENCE FUNCTION

The tastes and desires of an individual for economic goods at a given time may be described by ranking in the order of his preference all possible combinations of commodities and services. This ranking may be summarized by a function relating sets of commodities and services among which the individual is indifferent. Such a description is known as an indifference function.

[1] Although this paper is in all respects a joint product, and each author accepts full responsibility for all parts, the notions underlying Sec. II were originally due to W. A. Wallis and those underlying Sec. IV to M. Friedman.

The economist impounds in the indifference function all the psychological and sociological factors which determine consumer choices. The great merit of the indifference function is its ability to encompass these factors in their full generality and complexity and yet be free of irrelevant or erroneous assumptions about human psychology. Unlike the utility function, for example, it involves no assumption about the measurability of satisfaction. Those assumptions about tastes and preferences which are embodied in the indifference function are of real interest to the economist and are of a nature he is prepared to defend—if, indeed, he can find anyone prepared to attack them. The hypothesis that the indifference surfaces are negatively inclined rests on the fact that people generally prefer more of an economic good to less of it. The postulate that they are convex to the origin is based on the observation that willingness to exchange one good for a second—the rate of substitution which will leave the individual neither more nor less satisfied—increases with the abundance of the first good relative to the second; and it is reinforced by the consideration that surfaces not convex would lead to consumer behavior radically different from what is observed.

The indifference function has proved fruitful in theoretical economics because it states premises about consumer choices in a form that materially facilitates the correct deduction of rather intricate conclusions. Thus, it has improved our understanding of competitive and complementary relations among goods and of the interrelated effects of prices and income on quantities purchased, and it has led to important substantive results in the economics of welfare and in the theory of index numbers.

A second use to which the indifference function might be put is the organization and analysis of empirical data on consumer expenditures. Its fecundity in the first use is no guaranty against sterility in the second. The "natural rate of interest," "pure profit," "entrepreneurial expectations," and, for that matter, "general equilibrium," are all schizoid concepts—thoroughly competent in the field of deductive analysis but utterly incompetent for quantitative analysis. We consider here only the problem of giving quantitative expression to the indifference function and exclude consideration of its application in abstract form to problems of pure theory.

There are two fundamental forms of data from which the statistician might determine indifference surfaces. The first consists of a set of points known to lie on, or approximately on, the surfaces, together with information grouping these points according to the particular surfaces to which they correspond. That is to say, the data not only

specify a set of points but also indicate which subset of them represents one indifference surface, which a second, etc. The other basic type of data also specifies a set of points, but the information for co-ordinating them consists of the slopes of the indifference surface at each point. For either type of data appropriate statistical techniques for fitting a system of surfaces are readily available. [2]

There are also two approaches by which the observations might be obtained. Theoretically, either of these might yield data of either form; but it seems likely that in practice the first, which we call the "experimental" approach, will usually produce data of the first form and that the second, or "statistical," approach will ordinarily result in data of the second form.

II. THE EXPERIMENTAL APPROACH

The indifference function is defined for a single individual at a single time. Its purpose is to abstract from economics those psychological factors which, in conjunction with the cultural milieu, constitute the mechanism of choice. These two essential characteristics of the indifference function suggest that it can be evaluated most directly through the application of psychophysical experimental techniques to individual subjects. And in fact the first quantitative study of which we know did adopt such a procedure.

A suggestion by Professor Schultz that experimental methods be applied to this problem in economic theory led the eminent psychophysicist L. L. Thurstone to attempt a determination of the indifference function.[3] Thurstone starts with "five fundamental psychological postulates" which he states as follows:

1. Satisfaction increases with increase in the amount of the commodity possessed by an individual.
2. There is a lower limit in the amount of the commodity below which the owner will not or cannot barter.

[2] The first kind of data leads to a fairly familiar problem susceptible to straightforward least-squares treatment. Observations of the second type require a less familiar approach. From a mathematical function which is selected as the form of the indifference function, the marginal rates of substitution between all pairs of goods are obtained in terms of the unknown parameters and the quantities. Substitution of the observed marginal rates of substitution (ordinarily obtained from the ratios of the prices) and the associated quantities provides a set of "observation equations" from which the values of the parameters can be determined. In general, each observed point and the associated set of slopes results in one fewer observation equations than there are goods. For an excellent discussion of this problem for the special case of an indifference function which is a second-degree polynomial see A. Wald, "The Approximate Determination of Indifference Surfaces by Means of Engel Curves," *Econometrica*, VIII (1940), 144–75.

[3] L. L. Thurstone, "The Indifference Function," *Journal of Social Psychology*, II (1931), 139–67.

3. Motivation is defined quantitatively as the anticipated increment in satisfaction per unit increase in the commodity.
4. The motivation is finite when satisfaction is zero.
5. The motivation is inversely proportional to the amount already possessed.

The last postulate leads immediately to the deduction that the curves of indifference between pairs of commodities are hyperbolic. Since the general nature of the curves is one of the things the economist is most interested in determining empirically, it may seem unfortunate thus to assume the answer. Thurstone explains, however, that, "as a matter of fact, many other psychological postulates have been tried instead of the fifth postulate above, but the one that leads to Fechner's law seems to fit the experimental data better than any of the others that were tried." The fifth point seems, therefore, to be not a postulate at all but a conclusion from experimental data. There are many other alternative postulates, however, which also rationalize hyperbolic indifference curves, hence follow equally well from the data. Furthermore, as Thurstone points out, his data do not produce hyperbolas asymptotic to the axes, as would be expected from the fifth postulate; instead, they agree with the second (a contradiction of the fifth), which also is apparently not a true postulate but a conclusion from the experiment. The first, third, and fourth postulates relate to total and marginal utilities, concepts for which economists have often been criticized by psychologists—it is because the indifference function rids economics of these concepts that it has been so widely accepted—but fortunately these postulates had no bearing on the experimental procedures or results.[4] Thurstone's postulates, therefore do not restrict the interest of his experiment so drastically as might be supposed.

Three commodities—hats, shoes, and overcoats—were employed in the experiment, and indifference curves between all possible pairs were determined. A specified combination—for example, five of one item and three of another—was compared with a series of alternative amounts of the same two goods, the subject indicating in each comparison which combination he favored. To obtain the indifference curve, the preferred combinations were charted and a hyperbola asymptotic to the axes was fitted to their lower boundary. By repeating the procedure with different standard combinations other curves of the same system were obtained.

From the parameters of these hyperbolas Thurstone concludes: "For our subject the satisfaction curve rises fastest for overcoats

[4] The first postulate had a slight bearing, in that no combinations were compared in which both components of one exceeded both components of the other.

and slowest for hats . . . . . This is perhaps what one might expect from the common-sense values of hats and overcoats." He notes also that the data do not conform to hyperbolas asymptotic to the axes, but instead seem to be asymptotic to certain positive values, and concludes: "There is a lower limit for each commodity below which the subject is not willing to barter . . . . . This fact can be interpreted to mean that all barter ceases entirely when the quantity of a commodity reaches a lower limit which the subject evidently regards as a necessity."[5]

The economic significance of Thurstone's experiment is vitiated by a number of serious limitations. The only definition of the commodities in terms of which the subject made his hypothetical choices was that "the two alternative combinations would cost him the same and that the articles would be of his free selection within the general price range to which he is accustomed." It is questionable whether a subject in so artificial an experimental situation could know what choices he would make in an economic situation; not knowing, it is almost inevitable that he would, in entire good faith, systematize his answers in such a way as to produce plausible but spurious results. The results are consistent, for example, with the view that the subject's real criterion of choice, whether consciously or not, was related to the prices he customarily paid for the objects. Furthermore, the experiment did not attempt to handle rates of consumption but simply a single finite act of choice, a problem of slight relevance to economic theory. In addition, several flaws mar the statistical techniques, the most important of which is that certain conclusions are based on the estimated parameters of hyperbolas asymptotic to the axes, while others are based on the fact that the axes are not the asymptotes.

Although Thurstone's experiment might have been improved in detail, its more fundamental shortcomings probably cannot be overcome in any experiment involving economic stimuli and human beings. For a satisfactory experiment it is essential that the subject give actual reactions to actual stimuli. This requires conditions under which the reaction being studied is the only one on the basis of which the subject could produce systematic results; that is, any rationalizing scheme, conscious or unconscious, which he might adopt should have as its only possible datum the phenomenon under investigation. Questionnaires or other devices based on conjectural responses to hy-

[5] It is enlightening to contrast this interpretation of a portion of the curve parallel to an axis with the interpretations of economists. The latter usually regard it as representing not the irreducible minimum for the good to whose axis the curve is perpendicular but the saturation point for the good to whose axis the curve is parallel.

pothetical stimuli do not satisfy this requirement. The responses are valueless because the subject cannot know how he would react.

The reactions of people to variations in economic stimuli work themselves out through a process of successive approximation over a period of time. The initial response indicates only the first step in a trial-and-error adjustment; and unless the stimuli are within the immediate range of the subject's experience, this first step may depart widely from the ultimate response. If a realistic experimental situation were devised, it would, consequently, be necessary to wait a considerable time after the initial application of the stimulus before recording the reaction. Even an experiment of restricted scope would have to continue for so long a period that it would be exceedingly difficult to keep "other things the same."

Consider, for example, a hypothetical experiment designed to evaluate the indifference system of a child for candy and ice cream, two commodities that are consumed at fairly short intervals and for which fairly short durations of stimuli might therefore be expected to suffice. Suppose a dozen different combinations of candy and ice cream are offered to the child, and he is allowed to have whichever he chooses, a situation approximating that employed by Thurstone. It would clearly be necessary to offer the same combinations day after day, to give the subject an opportunity to experiment. Moreover, he could not be expected ever to settle on a single combination which he invariably chose. The attractiveness of variety would lead to alternations. At best, the rates of consumption per week, per month, or per year would eventually become approximately constant. When and if such constancy was reached, only a single observation would have been obtained. It would then be necessary to repeat the performance with different sets of combinations, in order to obtain additional observations. But would the observations obtained in this way relate to a single indifference function? Can other things be supposed the same over so long a period? Even if the subject were in an institutional environment characterized by an exceptional degree of stability, his needs and preferences would be affected by changes in the weather, in his physical and mental state, in the character of his meals, etc. If a long enough period were allowed for such factors to average out, the more fundamental changes in preferences associated with growth or even with continued exposure to the experimental stimuli would plague the investigation.

The fundamental difficulty is that economic phenomena are so integral a part of life that effective experimentation would require control of virtually the entire existence of the subject. Economic

stimuli likewise are not subject to control but are themselves continually evolving; and with the passage of time the response mechanism itself—the needs and tastes—undergoes changes which are to a large extent related to changes in the stimuli or even to the process of adjustment. These are more than technical or practical obstacles and indicate that it is probably not possible to design a satisfactory experiment for deriving indifference curves from economic stimuli applied to human beings.

There has been some very interesting economic experimentation on chimpanzees,[6] and it is not impossible (alas!) that some light on human tastes and preferences might be given by observation of the higher apes should they prove amenable to the experimental requirements of this kind of investigation. But for work on human beings the procedure which suggests itself, in view of the apparent impossibility of utilizing economic phenomena, is to formulate the problem in more abstract terms—to generalize it just sufficiently to render economic phenomena a subclass of a larger class of phenomena for which the same problem can be raised.

Needless to say, such generalization forfeits all the peculiarly economic aspects of the problem. And there is no assurance that problems of the same abstract form need be similar in solution. Nevertheless, the experiment proposed below might provide some information about the more general characteristics of indifference functions. We present it, therefore, not only as showing the extreme lengths to which one is extended in attempting to derive indifference functions experimentally but as a project that might be worth carrying out.

The essentials of the problem of quantifying the indifference function may be stated in the following form. A variety of measurable stimuli which are qualitatively different produce responses having a common element in terms of which ordinal comparisons may be made. The objective is to determine the relation between this response and the amounts of the stimuli. Since the response is measurable only ordinally, not metrically, this can be done by determining an implicit function connecting combinations of stimuli which yield the same response.

It is known that various colors produce differing size illusions, just as different commodities produce varying satisfactions. For example, a red cube with an edge of 2.45 inches will appear from a dis-

[6] See, e.g., John B. Wolfe, *Effectiveness of Token-Rewards for Chimpanzees* ("Comparative Psychology Monographs," Vol. XII, No. 5 [1936]); John T. Cowles, *Food-Tokens as Incentives for Learning by Chimpanzees* ("Comparative Psychology Monographs," Vol. XIV, No. 5 [1937]).

tance of about eight feet to be equal in size to a blue one with an edge of 2.49 inches.[7] This suggests a type of phenomenon which might be made the basis of an experiment. Two shades of purple can be produced by mixing red and blue in different proportions; squares of varying size in one shade could be compared with a square of fixed size in the other until the point of apparent equality is identified. If the red and blue are blended not by mixing the paints but by painting very small areas with the pure colors, in the manner of the Impressionist painters, they will still produce a single uniform hue if viewed from a distance. It will then be possible to measure the areas of red and blue in the two apparently equal squares. Thus, if a 41-centimeter square which is half blue and half red appears equal in size to a 40-centimeter square which is three-fourths red and one-fourth blue, it means that 840.5 square centimeters of red plus 840.5 square centimeters of blue produces the same size-response as 1,200 square centimeters of red plus 400 square centimeters of blue. Additional points on the same indifference curve would be found by using varying proportions between the colors in comparison with the same standard. Different curves in the same system would be secured by starting with a different size in the standard shade. Other systems could be obtained by using other colors.[8]

It should be noted in favor of this method that the combinations presented to the subject are blended into a homogeneous stimulus, so that he compares the two combinations as entities. Furthermore, it is possible to use any number of variables, thereby determining multidimensional indifference surfaces.

The particular values of the parameters would, presumably, have no relation to economic indifference functions; but for theoretical purposes it is ordinarily the general shapes which are important—for example in problems of complementarity, consistency (integrability), multiplicity of maxima, etc. There is also no empirical evidence that the shapes would bear enough resemblance to the desired economic functions, even in form, to be of value. Some presumptions on this score could, however, be established by experimenting with widely different types of phenomena. For example, sound, heat, weight, and

[7] W. Allen Wallis, "The Influence of Color on Apparent Size," *Journal of General Psychology*, XIII (1935), 193–99.

[8] As a matter of experimental technique, this procedure would be excessively laborious, but simpler ways of accomplishing the same end could be devised. For example, a color wheel would eliminate the extensive labor of painting small squares. Simply mixing the paints would be objectionable on the ground that the resulting fusion is not purely psychological but partly chemical; the colors produced by mixing pigments are not, in general, the same as those produced by mixing the same components on a color wheel.

other sensations might be used. It might also be possible to base experiments on variations in the conditions under which tests of manual or mental skills are applied. Similarity of patterns from a wide range of pheonomena would lend credence to the notion that the general form is essentially similar for economic pheonomena as well. Nevertheless, economic application of the results, even to very general problems in pure theory, will necessarily rest heavily on attenuant assumptions.

III. THE STATISTICAL APPROACH

According to the indifference function analysis of consumer behavior, the quantities of goods purchased define a point on the indifference surfaces at which the slopes are the ratios of the prices. This suggests the possibility of using data on consumer purchases for the quantitative determination of the indifference function. The obstacle to the approach is that the function is defined for a single person at a given time. It is obviously impossible to secure more than one observation at one time, whereas many observations covering a reasonably broad segment of the function are required.

If it can be assumed either that a given person has the same tastes at different times or that different persons have the same tastes at a given time, it will be possible to secure a number of observations relating to a single indifference function. Such assumptions, while never literally fulfilled, seem plausible. Certainly they seem more reasonable than those which have to be introduced in the experimental approach. They have the added merit of involving economic phenomena proper.

It seems reasonable to regard the tastes of an individual as essentially constant over fairly short periods. But the period of constancy would have to be fairly long to furnish data adequate for deriving indifference functions. The consumption of most goods should be measured in a time unit of not less than a year, in order to offset seasonal fluctuations in desires as well as the effects of anticipated seasonal changes in prices or income. For many goods, especially durable or semidurable goods, even two or three years would not be a long enough time unit. Thus, periods as long as five years would not yield more than five points for the determination of the indifference function.

The instances in which it might be valid to treat tastes as fixed over rather long periods are those in which there have been no significant alterations in the individual's economic milieu—his income and the price structure, in particular. It is just these instances, how-

ever, in which even data for a long period are statistically inadequate. If income, prices, and tastes are nearly constant, all the observed quantities and slopes fall in the neighborhood of a single point-slope. If, on the other hand, the individual has experienced a wide range of prices and incomes, his tastes have probably not remained constant, for past experience is surely one of the most important determinants of tastes at any given moment.

It must be remembered also that the sheer passage of time, accompanied by the transition from youth to middle age to senescence and by evolution in fashion, technology, and the general culture, effects alterations in preference. Furthermore, the adjustment of an individual to changed environment, even with fixed tastes, is so much a trial-and-error process that errors of adjustment may become almost biases.

A more promising procedure is the use of data on the purchases of different individuals at the same time. There are doubtless large groups of individuals who have similar tastes, so there should be no difficulty in securing enough observations, if these groups can be isolated by objective criteria.

Differences in the geographic and climatic characteristics and in the dominant cultural patterns of the various regions of the country and the diversity of modes of life in communities varying widely in size clearly make it necessary to use regional location and size of community as two of the criteria. But these criteria limit enormously the range of prices (and to some extent the range of incomes) within a group. The price of any commodity is presumably the same for all individuals in a community.

To eliminate the wide differences in taste which prevail even among individuals living in the same community, it would be necessary to isolate groups having similar social status, cultural and educational background, occupational attachment, and the like. Many if not most of these factors are closely associated with income; indeed, income would undoubtedly be the best single criterion of similarity of taste. Groups relatively homogeneous with respect to these taste-determining factors would, therefore, be relatively homogeneous with respect to income.

The use of data for the same time period but many individuals is thus no real solution. A group of individuals whose tastes could be supposed the same might vary somewhat, though not widely, in income status but would almost inevitably be confronted with similar price situations. Income variation alone is not sufficient to determine the indifference function.

There remains the possibility of combining the two types of data so far considered, namely, employing observations on many individuals for several periods of time. The prices would be the same for all individuals in each time unit but presumably, like the incomes of the individuals, would change over time.

This possibility is the only one that seems to offer any real hope of deriving indifference functions from market data, but it, too, is narrowly circumscribed. To avoid changes in tastes over time, only a few adjacent time units can be employed. If there are only slight variations in prices during this period, the statistical base of the indifference function will be narrow. Wide variations in the price level, on the other hand, ordinarily reflect such great changes in national income and in general social and economic conditions as to render untenable the assumption that tastes are constant. Thus, even for individuals with unchanged "real" income, a rapid decline in national income is likely to increase the desire for saving and decrease the desire for consumption and to alter the indifference curves for the separate consumption items. Moreover, it is hazardous to assume that individuals rapidly adjust their purchases to large and widespread changes; but, unless this is true, the price ratios and quantities do not represent point-slopes of an indifference function. (The trial-and-error nature of adjustments would, however, be far less serious than in the case of data for only one individual, because of the averaging-out of errors; in so far as there is similarity in the paths of adjustment—e.g., a lag—some "bias" would remain.) Material changes in the prices of a few goods only would be less inconsistent with the assumption of unaltered tastes and with fairly rapid adjustment to the price changes and might make possible a fairly satisfactory derivation of indifference surfaces for a subset of goods.[9]

While it is thus not entirely impossible to obtain indifference surfaces from market data, it seems highly unlikely that reliable results can be obtained for more than a small range of quantities for a few goods. The necessity of using data that reflect reactions to essentially the same indifference function implies a serious limitation on the degree of income and price variation that can be observed and hence in the scatter of points on which the indifference surfaces can be based. If these points cover a wide range, it is unlikely that they relate to the same system of indifference surfaces; if they cover a

[9] A possibility which may be mentioned is the use of data on total or per capita consumption for the country as a whole or subdivisions of it, together with average prices. This approach seems to be ruled out by the ambiguity of the concept of an indifference system for a collectivity or an "average" consumer.

narrow range, the indifference surfaces derived from them will be subject to wide margins of error and to statistical instability.

IV. INDIFFERENCE FUNCTIONS AS A FRAMEWORK FOR EMPIRICAL ANALYSIS

The difficulties of deriving quantitative counterparts of indifference functions reflect neither lack of data nor inadequacy of statistical technique. They are attributable to characteristics inherent in the logical structure of the indifference function.

Fundamental to the indifference curve analysis is a classification of the elements of consumer choice into three categories: (1) the objects among which choice is made—the *goods* that are represented by the axes of the indifference function; (2) the factors that determine which of the alternative combinations of goods open to an individual he will prefer—the *taste factors* that determine the character of an individual's system of preferences; and (3) the factors that determine what combinations of goods an individual is able to secure—the *opportunity factors* which restrict the range of combinations attainable by the individual.

This threefold classification seems straightforward and meaningful, and it has proved useful in problems of pure theory. But, as a framework for empirical analysis, does it really aid in filing the diverse and heterogeneous phenomena of consumer expenditures under headings that are distinctly different in analytical importance, or are cross-references so frequent and complex that the data might as well be left unfiled? A reconsideration of the difficulties discussed in Sections II and III in the light of this classification suggests the answer.

Consider, for example, size and composition of family, a factor which definitely affects consumption patterns (and which is almost always taken into account in family-budget studies, although it is neglected in theoretical analysis). From one point of view, family composition seems to be an opportunity factor, somewhat similar to income. A husband and wife with a given income, for example, are "better off" than a husband and wife with six children and the same income. From another point of view, however, family composition is a taste factor. Baby carriages, to cite an illustration, play a different role in the indifference systems of childless couples than in those of families with infants. Finally, family composition may even be treated as a good; for the satisfactions of an additional child are likely to be compared with the expenses incurred, and considerations of "price" may play a not inconsiderable role.

Regional location, another factor which family-budget studies usually recognize as determining spending habits, furnishes a second example. Region is clearly a taste factor: the relative evaluation of fur coats and bathing suits is not the same in California as in Wisconsin. But it is also an opportunity factor: the price of automobiles, for example, is higher in California than in Wisconsin. Nor can it be denied that region may also be a commodity: witness the number of retired individuals living in California and Florida, or the number of individuals who will consider similar positions in different regions equally attractive despite differences in salary.

Even so seemingly unambiguous an item as an automobile, which apparently is simply a good, raises problems. The "price" of a drive to the country, for example, is different for automobile owners than for others, and so is the price of items that can be bought more cheaply in the country than in the city, for example, farm produce.

Income is an opportunity factor par excellence, and yet it is probably the one thing most closely related to tastes. Prices also may have an important influence on tastes. Indeed, the opinion may be hazarded that while there are many items that belong unambiguously in the category of goods, there are few or none that may be classified as taste factors but not as opportunity factors, or as opportunity factors but not as taste factors.

A conceivable manner of handling these ambiguities is to include nearly all factors in the indifference function. For example, the fact that a family has a different set of preferences when it owns an automobile than when it does not is taken into account by the indifference function, in that the system corresponding with 0 on the automobile scale differs from that corresponding with 1. In the same way, eggs purchased by driving to the farm might be set up as one commodity and eggs purchased in the city as another. Similarly, such factors as region and family type could be taken into account as dimensions of the indifference function. The basic feature of indifference curve analysis that gives it importance, however, is that it seems to segregate the subjective factors that determine consumer behavior from the objective economic factors. The suggested solution—which in any event is quite impractical—would erase this distinction and leave a hodgepodge of psychology and economics.

The ambiguity of the classificatory criteria which are implicit in indifference curve analysis is, of course, the reason it is so difficult to specify reasonable conditions for deriving indifference curves from observational data. Satisfactory data can be obtained only if opportunity factors vary over a wide range while taste factors remain con-

stant; but this is clearly impossible because the opportunity factors and the taste factors are inextricably interwoven—are really the same factors under different aliases.

V. CONCLUSIONS

A quantitative statement of indifference functions would serve two purposes. First, it would provide more detailed premises about the form of the function than are now used in deductive analyses and would thereby lead to more specific conclusions. The discussion of Sections II and III holds out some hope of eventually delineating at least the broad outlines of the function, and these might be adequate or at any rate useful for this first purpose.

The second object of deriving indifference surfaces is to obtain exact knowledge of the quantitative relation of consumer expenditures to prices and income for the purpose of predicting the effect of changes in economic conditions on the consumption of various commodities and services. This second objective is far more important than the first; indeed, the first is not so much a different objective as a partial substitute for the second. According to the argument of Sections II, III, and IV, there is little prospect of material progress toward this second goal through the empirical derivation of indifference functions.

Recognition of the difficulties that block advances in this direction has not infrequently led to the conclusion that no sound progress toward understanding consumer demand is now possible. "Anything like significant income variations will lead to a changed 'standard of living' and a new indifference function," writes Stigler in a valuable survey of the limitations of statistical demand studies. "Until this problem is handled theoretically, I can see no significance in income elasticities based on anything but small income variations."[10] The data are adequate, the statistical techniques are satisfactory, but progress must await further advances in the theory of indifference functions. If this were so, the outlook would be gloomy indeed. For it is the contention of Section IV that further theoretical work on indifference functions cannot remove these obstacles to quantification; they are an inherent part of the theory and represent not uncharted territory but seas in which no solid ground for empirical work exists.

These obstacles, fortunately, are inherent not in the problem but merely in the approach through indifference functions. Indeed, an investigator with his gaze firmly fixed on the objective of relating consumer expenditures to prices, income, and tastes or needs would prob-

[10] George J. Stigler, "The Limitations of Statistical Demand Curves," *Journal of the American Statistical Association,* XXXIV (1939), 469–81.

ably not seriously consider using the indifference function as an intermediary. Only if he were steeped in theoretical analysis to the point of oblivion to the real problem and the data available for its solution would he mistake the scaffolding set up to facilitate logical analysis for the skeletal structure of the problem.

In point of fact, empirical workers have adopted the direct approach of isolating factors correlative with consumer demand and measuring the relationships. Such factors as income, wealth, prices, family type, occupation, age, nationality, regional location, type of community, ownership of home or automobile, etc., have been studied with a view to determining which are most intimately associated with spending patterns, and in what manner. Whether these are taste factors or opportunity factors is irrelevant. If, for example, the food expenditures of families of a certain group uniformly increase by 45 per cent when income increases from $1,000 to $2,000 and decrease correspondingly when income declines from $2,000 to $1,000, what does it matter whether the changes reflect responses to fixed indifference functions or a concomitant alteration of tastes? The important issues in interpreting studies of this type are the extent to which the adjustment of given families to changes in circumstances can be inferred by comparing the contemporaneous behavior of families in different circumstances, and the extent to which the observed relations are stable through time. These questions are subject to direct empirical investigation.

Largely because so much attention has been devoted to indifference analysis, theoretical work in the field of consumer expenditures has lagged behind the statistical work. There is much to be gained by concentrating some heavy theoretical artillery on the logical structure implicit in the practical work. What, for example, is a sound mathematical formulation of the income-expenditure curve? What definitions and groupings of goods are most relevant for general economic analysis? How are quality differences in goods to be treated? Are there sound procedures for quantifying such variables as family type, for example, by the construction of so-called "ammain" (adult-male-maintenance) units?

By way of summary, let us repeat that we fully recognize the power of the indifference function in pure theory (though the considerations raised in Sec. IV suggest that this power is definitely limited). We doubt, however, that it has any material value for the organization of empirical data. Fortunately, there are superior and more direct routes to the ends which its quantification would fulfil. These routes have heretofore been the domain chiefly of statisticians; they require and are worthy of more intensive theoretical analysis.

EXPENDITURE PATTERNS OF FAMILIES OF DIFFERENT TYPES

R. G. D. ALLEN
London School of Economics

SUPPOSE we have a group of families homogeneous as regards size, age and sex composition, social position, and area of residence. The way in which a family of the group plans its expenditure depends on the amount of income available in a given period and on the variation of the individual tastes of the family from the "norm" of the group. For the range of incomes shown by the set of families we can obtain a regression equation relating expenditure on an item of consumption to total income. The equation exhibits the effect income has upon this item of expenditure for a "typical" member of the group; we have eliminated variations of tastes among individual families, these variations being indicated by differences between actual expenditures and those given by the regression equation. If n items of expenditure appear in the complete family budget, then there is a set of n regression equations which serves to describe completely the "income effect" characteristic of the group of families and which can be termed the "expenditure pattern" of the group.[1]

In the simplest case the expenditure pattern is composed of linear regression equations, that is, the expenditure (x_r dollars per year) by the "typical" family on the rth item is related to the income (x dollars per year) of the family by

$$x_r = a_r + b_r x , \qquad (r = 1 , 2 , \cdots , n)$$

where a_r and $b_r (r = 1 , 2 , \cdots , n)$ are constants for the group of families. The regression equations must be obtained subject to the control

$$\sum_{r=1}^{n} a_r = 0 \quad \text{and} \quad \sum_{r=1}^{n} b_r = 1 .$$

We shall assume that all b_r's are positive.[2]

[1] An analysis of the variations of individual expenditures about the "typical" regression relations would serve to indicate how significant and well established are these relations (see R. G. D. Allen and A. L. Bowley, *Family Expenditure* [London, 1935], pp. 59 ff). This stage in the analysis is not carried out in the present study.

[2] For a few particular items (e.g., flour or margarine) it is possible that negative values of b_r are obtained (see Allen and Bowley, *op. cit.*, p. 41). We can eliminate this case, however, by a suitable grouping of items.

The elasticity of expenditure on the rth item at the income level x is

$$\eta_r = \frac{x}{x_r}\frac{d\,x_r}{d\,x} = \frac{1}{1+\dfrac{a_r}{b_r x}}.$$

A scale of urgency among different expenditures at any given income level is fixed by the order of the values of $\eta_r (r = 1, 2, \cdots, n)$. When η_r is less than unity, we have a necessary item of expenditure, and the smaller the value η_r, the more urgent is the expenditure. Luxury items are similarly characterized by values of η_r greater than unity.[3] It is seen at once that, if η_r is less (greater) than unity for one income level, it is less (greater) than unity for all income levels and increases (decreases) toward unity as income increases indefinitely. The linear expenditure pattern has, in fact, the property that the scale of urgency (fixed by the spread of values of η_r for $r = 1, 2, \cdots, n$) becomes narrower as income rises; as needs are increasingly satisfied, necessary items become less urgent in the budget while other expenditures lose some of their luxury nature.

A necessary expenditure ($\eta_r < 1$) has a_r positive, and purchases of the item are made no matter how low the income. The larger the value of a_r, and the smaller the value of b_r (indicating the rate of growth of the expenditure as income rises), the more urgent is the item. The expenditure on a marked necessity is large even at low income levels, and it is little affected by increases in income. The proportion of income spent on the rth item is

$$p_r = \frac{a_r}{x} + b_r.$$

An urgent expenditure, therefore, forms a decreasing proportion of income as income rises. A luxury expenditure ($\eta_r > 1$) has a_r negative, so that purchases begin only at a definite income level and form an increasing proportion of income as the level rises.

The expenditure pattern of a homogeneous group of families exhibits the effect a variable income has on the expenditures typical of the group. The patterns characteristic of different homogeneous groups can be compared, for example, by means of the set of constants a_r and $b_r (r = 1, 2, \cdots, n)$ and of the values of η_r and $p_r (r = 1, 2, \cdots, n)$ at various income levels. The comparison may be (a) between

[3] See Allen and Bowley, *op. cit.*, pp. 12–13.

families of different social positions (e.g., wage-earner, clerical or professional families); (*b*) between families in various regions, in large and small cities or in urban and rural communities; or (*c*) between families of varying size or age and sex composition. Relatively little has, however, been done in this direction; indeed, the scope of practically all family-budget investigations has been too limited for such comparisons to be possible. The influence of the factors of social position, location, and family size on modes of expenditure thus remains obscure.

The effect of the size and composition of families on the characteristic pattern of expenditure is of particular interest. Instead of allowing family budgets to "speak for themselves" in a way such as that described above, investigators have attempted to eliminate the influence of varying family composition by the use of "equivalent adult" scales settled in advance. This approach to the problem is of more than doubtful utility; the ways in which families actually spend their incomes may have little connection with the manner in which they "should" do so according to dietary principles.

An unequaled opportunity for an analysis of expenditure patterns typical of different groups of families is provided by the Urban Study of Consumer Purchases (1935–36) conducted by the Bureau of Labor Statistics (U.S. Department of Labor) in co-operation with other governmental agencies.[4] This study was based on a strictly controlled sample of families. Expenditure schedules, relating to a whole year, were obtained from a sufficient number of families in each urban community selected for investigation, in each of several occupational strata, in each type of family composition distinguished and in each of a series of income brackets. Our main object here is to isolate the factor of family size and composition and we are able to separate six family types as follows:

Type 1—Man and wife only
Type 2—Man, wife, and one child
Type 3—Man, wife, and two children
Type 4—Man, wife, and three or four children
Type 5—Man, wife, and five or six other persons (one at least being a child)
Type 6—Man, wife, one other adult, and perhaps one other person

[4] A number of papers and bulletins have been published on the methods, objectives, and findings of the Urban Study. See Erika H. Schoenberg and Mildred Parten, "Methods and Problems of Sampling Presented by the Urban Study of Consumer Purchases," *Journal of the American Statistical Association*, XXXII, 311–22; A. D. H. Kaplan, "Expenditure Patterns of Urban Families," *Journal of the American Statistical Association*, XXXIII, 81–100; U.S. Department of Labor, *Family Income and Expenditure in Chicago, 1935–36* (Bureau of Labor Statistics Bull. 642 [2 vols.; Washington, 1939]).

Families are included only if they contain husband and wife, both native white, if they have not received relief during the year investigated, and if they satisfy certain minor "eligibility" conditions.[5] Type 1, though comprising the smallest families, is not perfectly homogeneous; young married people and older couples whose children have grown up and left home are alike included. Types 2, 3, and 4 are reasonably homogeneous, covering families with varying numbers of young children. Type 5 represents the very large family, usually containing children under and over sixteen years of age and possibly other adults as well. Families of Type 6 are of middling size and largely of adult composition.

To separate off other factors which may affect family expenditure, we analyze the schedules of families of our six types in otherwise fairly homogeneous groups. Wage-earner families[6] have been selected in cities of varying size in one relatively small region and arranged in three groups—families in the metropolis of Chicago, in the large city of Columbus (Ohio), and in the three smaller cities of Muncie (Ind.), New Castle (Pa.), and Springfield (Ill.). Two groups of clerical families[7] have also been taken, in Chicago and in Columbus (Ohio), respectively. We have, in this way, five separate collections of schedules for our analysis of the effect of family size on modes of expenditure. We are also able to distinguish some differences between the expenditure patterns characteristic of wage-earner and clerical families and of families in larger and smaller communities.

For the purposes of the present broad analysis, we have taken only six categories of expenditure, together making up the complete family budget:

Food—including meals away from home, candy, and drinks

Housing, fuel and light—including an imputed value when accommodation is owned or obtained free

Recreation and personal care—comprising all amusements, sport, tobacco, newspapers and other reading, hairdressing and all toilet articles

Clothing—personal items only

Transportation and furnishings—comprising all forms of (nonbusiness) travel, purchase and upkeep of automobiles, and all household operation expenses other than fuel and light

[5] See U.S. Department of Labor, *op cit.*, II, 208. The numbering of the family types given here differs from that adopted in this bulletin. A child means a boy or girl under the age of sixteen years.

[6] The occupational grouping of families is determined by the source from which the bulk of the family income is drawn.

[7] Clerical workers comprise those in offices who are not salaried executives, together with such employees as salesmen. The sample of clerical families in the three smaller cities was too small to yield significant results and has been omitted.

Miscellaneous and savings—including medical and educational expenses, gifts, taxes, and net savings

It can be noted that, of the items apart from food, clothing, recreation and personal care are almost entirely *personal* categories, while housing, fuel and light, transportation and furnishings are predominately *household* items of expenditure.

The expenditure schedules number about 100 of each family type in each of the five collections analyzed, a total of approximately 3,000 schedules. The families of each type are distributed fairly evenly over about ten income brackets,[8] the over-all ranges being as follows:

INCOME RANGE (DOLLARS PER YEAR)

	Chicago	Columbus	Smaller Cities
Wage-earner families:			
Types 1, 2, 3	500–3,000	500–3,000	250–2,500
Types 4, 5, 6	500–5,000	500–5,000	250–2,500
Clerical families	750–5,000	750–5,000	....

Linear regression equations, relating yearly expenditures in each of the six categories to total income per year, have been fitted graphically for each group of families, points being first plotted from the average expenditures found in each income bracket.[9] Since each regression line was deduced from data relating to only about 100 families, the plotted average points were found to vary a good deal about the fitted lines. The variation, however, appeared to be largely random, and little evidence of any definite curvilinear relation was noticed. It can be maintained that the linear relations provide good descriptions of the connections between expenditure and income for typical families of the types distinguished, at least as far as the broad categories of expenditure are concerned.

An expenditure pattern comprising six regression lines is isolated, in this way, for each type of family in the three collections of wage-earner, and in the two collections of clerical, schedules. A great number of comparisons between these patterns is now possible. The comparisons can be made from the constants (a and b) of the regression lines and also from the values of the expenditure elasticities (η) and of the proportions (p) of expenditure to total income. It is

[8] The original collections included a few schedules for small wage-earner families with higher incomes than here specified. These families have been omitted as being unusual.

[9] The sample, being controlled by income, provided approximately equal numbers of families in each income bracket. The graphical fitting was done subject to the conditions that the sum of the intercepts of the lines (corresponding to the six categories of expenditure) must equal zero and the sum of the slopes must equal unity. No attempt was made to fit lines by any "least-squares" method.

to be remembered, however, that η and p vary with the level of income. To provide figures easily tabled, we take the values, $\bar{\eta}$ and $\bar{p}$, at the arithmetic average income of the family type concerned.[10] The complete figures for the thirty expenditure patterns, obtained for six family types in each of our five separate groups of schedules, are set out in the tables in the appendix below. The regression lines themselves are easily reconstructed from the constants a and b in each case.

As a preliminary, we may look at the average incomes of groups of families of various types and at the proportions of these incomes devoted to the six categories of expenditure distinguished here. When the earnings of the husband constitute almost the only source of income, as is the case of families of Types 1, 2, 3, and 4, the recorded average income differs little from one type of family to another, though the average naturally varies considerably as between wage-earner and clerical families and in communities of different sizes. Subsidiary earnings increase in importance when grownup children or other adults are included in the family (Types 5 and 6), and the average income is then markedly higher. The proportion of average income spent on food is lowest, as we should expect, for the man-and-wife families and then increases quite rapidly as the size of the family grows, by the addition of children (Types 2–5). The very large families (of Type 5) spend, at their average income, a proportion on food which is about one-third greater than that spent by the small families. Families of adult composition (of Type 6), like the families comprising only husband and wife, expend a relatively small proportion of their average income on food. A large part of average income is devoted to housing expenses by the smallest families, while the proportion is less for families with young children and lower still for the very large family groups. The amount spent on rent, fuel, and light, in fact, tends to vary *inversely* with housing needs. Recreation and personal care account for a surprisingly constant percentage of average income, being generally rather less than 8 per cent for all types and groups of families investigated here. Personal care alone comprises a steady 2 per cent of income, while recreation varies a little between 5 and 6 per cent. Except in the cases of the largest families, clothing expenses in proportion to average income also vary little from one family type to another, about 8 per cent, being recorded for wage-earner families and a rather higher proportion for the

[10] The average income is derived not from the controlled sample but from a random sample giving frequency distributions of families by income.

clerical groups. It does appear, however, that the very large families spend considerably more of their higher average incomes on clothing, the proportion exceeding 10 per cent in all cases examined. Expenditure on transportation and household furnishings forms a high percentage of income for small families, and the relative importance of this item generally falls markedly as the size of family increases. Savings and miscellaneous expenses are very much higher for the "adult" families (of Types 1 and 6) than for the families with young children.

Certain broad differences between wage-earner and clerical families and between families in the larger and smaller cities are also apparent. Wage-earner families, with lower average incomes, find it necessary to devote larger proportions to food than clerical families; as a result, their miscellaneous expenses and savings are at much lower levels. It is remarkable, however, that expenditures on transportation and furnishings by wage-earner families form nearly the same proportions of average income as are found for the clerical families. Families in Columbus and the smaller cities have lower average incomes than the corresponding families in Chicago, and their typical modes of expenditure are different. Not only do they spend less on housing but their food expenditures also form considerably smaller proportions of average income. On the other hand, the amounts devoted by families in the smaller communities to transportation, furnishings, and miscellaneous items are larger in relation to their incomes.

Passing to a detailed examination of expenditure patterns, we notice first that the purely personal categories of expenses display a very striking property. The constant term in the regression equation for expenditure on recreation and personal care does not differ significantly from zero in any case; and, as a result, the value of elasticity of expenditure varies little from unity. This category of expense, therefore, forms an almost constant proportion of income *at all income levels.* The proportion (rather less than 8 per cent) that the joint expenditure on recreation and personal care bears to total income is little influenced by differences of income, of family size and composition, of social position, or of size of community. As income rises, the total family expense on these items increases in rough proportion and in approximately the same proportion, moreover, for all types and groups of families. The remaining personal expenditure, on clothing, is generally something of a luxury item; the intercepts of the regression lines, though small, are generally negative and the values of η are usually somewhat greater than unity. The proportion

of income spent on clothing thus rises, though only at a slow rate, as family income increases. This is true, with only small quantitative differences, of almost all types and groups of families. The only notable variation occurs in the case of the largest families (of Type 5); here clothing expense is definitely a luxury item in the family budget, showing a more rapid increase, both absolutely and in proportion to income, as income rises. It is surprising that the mode of expenditure on clothing by clerical families does not differ significantly from that by the wage-earners. At a given income level, the two groups of families spend approximately equal amounts on clothing.[11]

The patterns of expenditure need more careful interpretation in respect to other categories of expenses—on food, rent, and household purchases, miscellaneous items, and savings. The patterns are found to vary considerably from one family type to another and also to differ according to the social position and location of the families. We find it convenient to take wage-earner families first, analyzing the patterns displayed by the different family types, from the smaller and simpler families to those of larger size and more complex composition. Any significant differences shown by clerical families can then be noticed, together with such broad variations between the various communities as may appear.

For wage-earner families comprising man and wife only, food and housing expenses are, as we should expect, very urgent items in the family budget. It is surprising, however, that housing expense is the more urgent category of the two. Expenditure on rent, fuel, and light increases quite slowly as income rises while that on food expands with some rapidity. At the average income level the elasticity of expenditure on housing is generally less than 0.5, while that on food is significantly greater. House furnishings and (particularly) transportation are items of expense of definitely luxury nature. Few families with low incomes run automobiles, and little is spent on any form of transport. As income increases, however, more families are found owning automobiles, and the total amount expended on transport increases very rapidly. But the main "luxury" category of expenditure is the group of miscellaneous expenses together with net savings. There are net deficits at low levels and substantial savings at high levels of income; debts incurred exceed miscellaneous expenses (including gifts and taxes) at all incomes, on the average, up to $1,000 per year in Chicago, up to $830 per year in Columbus, and up to $720

[11] Clerical families expend a larger proportion of their *average* income on clothing than wage-earner families simply because this average income is higher.

per year in the small cities. Net savings, of course, appear only in much higher income brackets.

Passing to wage-earner families of Type 2, containing one young child in addition to the husband and wife, we find rather different modes of expenditure. Housing expense, in particular, is a much less urgent item for these families than for those without children; less is spent on rent, fuel, and light out of low incomes while expenditure increases quite rapidly as income rises. On the other hand, the increase in luxury expenditure (transportation, furnishings, miscellaneous, and savings) which follows a rise in income is generally less marked here than for the simpler families. Net savings, for example, are seldom considerable even in high-income brackets.

The effect that large numbers of children have on the modes of expenditure of wage-earner families can be traced by comparing the patterns obtained for one-child families (of Type 2), first with those for the rather larger families (of Types 3 and 4), and then with those for the very large families (of Type 5). As the number of children increases, the expenditure line for food tends to shift upward and to become steeper; a higher general level of food expense is necessary and there is more response to rises in income. The main change in the mode of expenditure on housing, at least in the larger cities, is to be found when we come to the very large families; rent, fuel, and light form a more urgent category for these families than for the smaller ones, expenditure being large even at low incomes and affected relatively little as income rises. The pressure of housing need is very marked for the large families. Expenditures on transportation and furnishings show a definite tendency to become less responsive to income changes as we pass to the larger families. When there are many children in the family, these expenses never make up a large proportion of total income, and they increase quite slowly as income grows. The effect of increasing size of family on miscellaneous expenses and net savings appears to be somewhat erratic and generally lacking in direction.

A final comparison can be made between wage-earner families of the simple two-person type and those of the larger adult composition (Type 6). Housing expense is a more urgent category of expenditure than food purchases for both groups of families. The difference is, however, more marked for the larger families which show, in general, steeper food expenditure lines and flatter housing expense lines than do the simpler families. The family with several adult members needs extensive housing accommodation and spends heavily on rent, fuel,

and light even when family income is low, while increases in income have relatively little effect. Expenditure on transport and house furnishings by adult families of Type 6 is much less of a luxury item in the budget than for the families comprising only man and wife; a substantial proportion of total income is devoted to these expenses at all levels of income. On the other hand, the expenditure pattern of the larger adult families is characterized by a miscellaneous and savings expense line of marked steepness and with a large negative intercept. Though such families can show considerable net savings at high income levels, they are involved in large deficits in the lower, and even in the average, income brackets. For example, for wage-earner families of Type 6 in Chicago, debts incurred are, on the average, in excess of miscellaneous expenses (including gifts and taxes) for incomes up to nearly $1,500 per year.

The expenditure patterns of clerical families can now be compared, family type by family type, with those of the wage-earner families. Perhaps the most striking result is the absence of large and systematic differences between the patterns of the families at the different social levels. For families comprising man, wife, and one young child, indeed, the clerical expenditure pattern is almost identical with the wage-earner pattern. As the size of family increases, certain differences become apparent. The food expenditure line, for example, tends to be flatter for clerical families; the latter prefer to devote increment in income more to other items in the budget. It is to be noticed that the larger clerical families (particularly of Types 5 and 6) seem to utilize higher incomes in the satisfaction of housing needs to a greater extent than the wage-earner families; the housing expense lines are generally higher and steeper for the large clerical families than for the corresponding wage-earner families. The mode of expenditure on transportation also differs as between the smaller wage-earner and clerical families. This item is much less of a luxury for the clerical families, which appear to run automobiles and to spend quite extensively on other forms of transport even at low levels of income.

The results so far described apply, broadly speaking, to families in each of the communities investigated. Certain differences, of a relatively minor nature, are to be noticed between the expenditure patterns of families in Chicago and in the smaller cities. It appears, for example, that expenditure on the urgent items (food and housing) by small families in Chicago is more responsive to income changes than expenditure by similar families in the other communities. Again,

the general level of net deficits is considerably greater among families with children (notably of Types 3 and 4) in Chicago than elsewhere, and net savings seldom appear even when incomes are considerable.

We may conclude that personal expenses, particularly on recreation and personal care, are governed surprisingly little by size of family or by any of the other factors considered. These expenditures vary nearly in proportion to income, and roughly the same proportion holds for all types of families. A large family spends as much on personal items in the aggregate, and hence proportionately less per person, as does a small family with the same total income. On the other hand, the urgent items (food, housing, fuel, and light) and the luxury categories (transportation, furnishings, miscellaneous expenses and savings) appear in the expenditure patterns in different ways for families of varying size and composition. The differences attributable to the factor of family size are certainly as marked, and probably more marked, than any variations that appear between wage-earner and clerical families or between families in the larger and the smaller cities. We should certainly be in error in ignoring differences in the size and composition of families in any analysis of patterns of expenditure of families at different social levels in various communities.

It is often maintained that families of varying size and composition, but otherwise homogeneous as regards social position and other factors, have essentially the same modes of expenditure. All that is necessary, it is said, is to reduce income and each expenditure to an "equivalent adult" basis. The results we have set out throw some light on this point and enable us to see to what extent "equivalent adult" scales are capable of eliminating the effect of changes in family size.

Assume that "equivalent adult" scales have been devised for each category of expenditure considered and let n_r be the number of equivalent adults, in a family of a certain size and composition, for expenditure on the rth item. The n_r's vary from one item to another since the "equivalent adult" scale for (say) food is different from that for housing. Let n be the number of equivalent adults in the same family for all expenditures (and for income); that is, n is some average of the various n_r's for the family. Assume, further, that the relation between any expenditure and income is of the linear form discussed above. Then, if the "equivalent adult" method of eliminating the effect of varying family size is valid, a single linear relation holds between expenditure and income, each per equivalent adult,

$$\frac{x_r}{n_r} = A_r + B_r \frac{x}{n},$$

where A_r and B_r are constants for families *of all sizes.* So

$$x_r = (A_r n_r) + (B_r \frac{n_r}{n}) x.$$

Fitting a regression line to expenditure of families of a given size and composition only, we should get an intercept proportional to n_r and a slope proportional to n_r/n. For food expenditure or a personal item such as clothing, n_r should increase more rapidly than n as the size of the family grows. Hence, with increasing family size, the intercepts of the different regression lines should increase rapidly in numerical magnitude while the slopes should also increase, but more slowly. On the other hand, for a household item such as rent or furnishings expense, n_r increases less rapidly than n, the intercepts should increase fairly slowly in numerical magnitude and the slopes should decrease. In all cases, since the intercepts increase numerically, we expect that the "spread" of the scale of urgency should increase as the size of the family grows; that is, urgent items tend to become more urgent and a luxury item more of a luxury.

These conclusions are not confirmed by the results of our analysis. In the first place, we have found that expenditures on such personal items as recreation and clothing are influenced scarcely at all by differences of family size and certainly not in the way the above theory indicates. For other categories of expenditure, there are particularly marked discrepancies between theory and practice for an increase in family size from Type 1 to Type 2 (corresponding to the addition of a child) and from Type 1 to Type 6 (when at least one adult is added). The "spread" of the scale of urgency, for example, is generally found to diminish or to remain unchanged, instead of increasing, as we pass from families of Type 1 to either of the larger family groups. For families with children (of Types 2, 3, and 4) our results are perhaps somewhat more in line with the theory. But, even here, the appropriate equivalent scales would appear to be different for families of varying social position and for the first or second as compared with later children. Children in clerical families, for example, seem to have greater weight in the food scale than those in wage-earner families, while the first or second child usually has greater weight than the third or fourth.

The present evidence tends to suggest, therefore, that "equivalent adult" scales are not capable of eliminating completely the effect on

modes of expenditure of varying family composition. The failure appears particularly as between families comprising husband, wife, and children, on the one hand, and families with adult members (apart from the man and wife), on the other. It seems just possible that scales could be devised to allow roughly for changes in family size due *solely* to varying numbers of young children in the family. To do even this, the scales may have to take account of differences between various kinds of personal and household expenses, between occupational groups of families and between the addition of the first child and the addition of further children. Equivalent adult scales, it would appear, must be used with caution and in a limited field.

APPENDIX

The following tables show expenditure in relation to income for native-white families, not on relief, in certain East Central cities for 1935–36. In these tables a represents the intercept and b the slope of the regression line of expenditures on income, and $\bar{\eta}$ is the elasticity of expenditure and $\bar{p}$ the expenditure as a proportion of income, each at the average income of the group of families concerned.

I. CHICAGO: WAGE-EARNER FAMILIES

EXPENDITURE ON:

	Food	Housing, Fuel, and Light	Recreation and Personal Care	Clothing	Transportation and Furnishings	Miscellaneous and Savings
Family Type 1. Average Income: $1,471 per Year						
a	195	215	0	—20	—130	—260
b	.190	.140	.074	.097	.233	.266
$\bar{\eta}$	0.59	0.49	1.00	1.16	1.61	2.98
$\bar{p}$	.323	.286	.074	.083	.145	.089
Family Type 2. Average Income: $1,475 per Year						
a	190	150	7	—15	—130	—200
b	.225	.167	.073	.091	.215	.229
$\bar{\eta}$	0.64	0.62	0.94	1.13	1.69	2.45
$\bar{p}$	.354	.269	.078	.081	.127	.094
Family Type 3. Average Income: $1,526 per Year						
a	220	145	4	0	—110	—260
b	.248	.167	.073	.078	.216	.218
$\bar{\eta}$	0.63	0.64	0.97	1.00	1.50	4.64
$\bar{p}$	.392	.262	.076	.078	.144	.047
Family Type 4. Average Income: $1,567 per Year						
a	265	175	—5	—40	—60	—335
b	.225	.163	.084	.110	.152	.273
$\bar{\eta}$	0.57	0.59	1.04	1.30	1.34	4.61
$\bar{p}$	.394	.275	.074	.084	.114	.059
Family Type 5. Average Income: $2,049 per Year						
a	270	245	10	—5	—30	—490
b	.272	.106	.074	.108	.119	.321
$\bar{\eta}$	0.67	0.47	0.94	1.02	1.14	3.92
$\bar{p}$	.404	.226	.078	.106	.105	.082
Family Type 6. Average Income: $1,894 per Year						
a	290	325	0	—15	—35	—565
b	.194	.084	.076	.089	.160	.397
$\bar{\eta}$	0.56	0.33	1.00	1.10	1.13	4.02
$\bar{p}$	.347	.256	.076	.081	.143	.099

II. COLUMBUS (OHIO): WAGE-EARNER FAMILIES

EXPENDITURE ON:

	Food	Housing, Fuel, and Light	Recreation and Personal Care	Clothing	Transportation and Furnishings	Miscellaneous and Savings
	Family Type 1. Average Income: $1,441 per Year					
a	205	260	11	—31	—190	—255
b	.137	.092	.059	.092	.314	.306
$\bar{\eta}$	0.49	0.34	0.87	1.31	1.73	2.37
$\bar{p}$	.278	.274	.068	.070	.181	.129
	Family Type 2. Average Income: $1,409 per Year					
a	195	175	—5	—36	—55	—275
b	.159	.133	.082	.105	.213	.308
$\bar{\eta}$	0.53	0.52	1.05	1.33	1.22	2.70
$\bar{p}$	.298	.256	.078	.079	.175	.114
	Family Type 3. Average Income: $1,437 per Year					
a	155	165	32	—31	—70	—250
b	.228	.132	.064	.100	.206	.270
$\bar{\eta}$	0.68	0.53	0.74	1.28	1.30	2.84
$\bar{p}$	.335	.247	.086	.078	.158	.095
	Family Type 4. Average Income: $1,397 per Year					
a	230	145	10	—4	—85	—295
b	.202	.149	.073	.083	.200	.293
$\bar{\eta}$	0.55	0.59	0.91	1.04	1.44	3.57
$\bar{p}$	.365	.253	.080	.080	.139	.082
	Family Type 5. Average Income: $1,818 per Year					
a	210	180	—22	—100	—125	—145
b	.255	.119	.092	.169	.199	.166
$\bar{\eta}$	0.69	0.54	1.15	1.47	1.52	1.95
$\bar{p}$	.370	.219	.080	.115	.131	.085
	Family Type 6. Average Income: $1,858 per Year					
a	200	250	—3	—11	—45	—390
b	.196	.095	.080	.092	.183	.354
$\bar{\eta}$	0.64	0.41	1.03	1.07	1.16	2.46
$\bar{p}$	.305	.229	.078	.086	.158	.144

III. MUNICE (IND.), NEW CASTLE (PA.), AND SPRINGFIELD (ILL.): WAGE-EARNER FAMILIES

EXPENDITURE ON:

	Food	Housing, Fuel, and Light	Recreation and Personal Care	Clothing	Transportation and Furnishings	Miscellaneous and Savings
Family Type 1 Average Income: $1,247 per Year						
a	145	185	0	—40	—65	—225
b	.168	.129	.070	.105	.215	.313
$\bar{\eta}$	0.59	0.46	1.00	1.48	1.33	2.34
$\bar{p}$	.283	.279	.070	.071	.162	.134
Family Type 2. Average Income: $1,275 per Year						
a	220	135	11	—7	—125	—235
b	.138	.148	.070	.084	.271	.289
$\bar{\eta}$	0.44	0.58	0.89	1.06	1.58	2.75
$\bar{p}$	.311	.255	.079	.079	.171	.105
Family Type 3. Average Income: $1,284 per Year						
a	235	130	10	—5	—20	—350
b	.150	.155	.067	.082	.166	.380
$\bar{\eta}$	0.45	0.61	0.89	1.05	1.09	3.52
$\bar{p}$	.333	.254	.075	.078	.152	.108
Family Type 4. Average Income: $1,301 per Year						
a	220	160	13	19	—105	—305
b	.216	.126	.063	.075	.241	.280
$\bar{\eta}$	0.56	0.50	0.86	0.83	1.53	6.09
$\bar{p}$	.384	.250	.073	.090	.158	.046
Family Type 5. Average Income: $1,426 per Year						
a	240	115	—12	—21	—50	—275
b	.245	.148	.082	.122	.152	.251
$\bar{\eta}$	0.59	0.64	1.11	1.14	1.29	4.33
$\bar{p}$	.413	.230	.074	.107	.118	.058
Family Type 6. Average Income: $1,423 per Year						
a	195	230	13	—5	—65	—370
b	.190	.107	.057	.088	.168	.390
$\bar{\eta}$	0.58	0.40	0.86	1.05	1.38	3.02
$\bar{p}$	.328	.270	.066	.084	.122	.129

IV. CHICAGO: CLERICAL FAMILIES

EXPENDITURE ON:

	Food	Housing, Fuel, and Light	Recreation and Personal Care	Clothing	Transportation and Furnishings	Miscellaneous and Savings
Family Type 1. Average Income: \$1,912 per Year						
a	270	315	20	—10	—70	—525
b	.133	.096	.068	.099	.190	.414
$\bar{\eta}$	0.49	0.37	0.87	1.06	1.24	2.97
$\bar{p}$	.274	.261	.078	.094	.153	.140
Family Type 2. Average Income: \$1,890 per Year						
a	185	195	14	—25	—112	—257
b	.220	.150	.068	.099	.197	.266
$\bar{\eta}$	0.69	0.59	0.90	1.16	1.43	2.04
$\bar{p}$	.318	.253	.075	.086	.138	.130
Family Type 3. Average Income: \$2,021 per Year						
a	285	200	—12	—45	—125	—305
b	.196	.150	.089	.117	.205	.243
$\bar{\eta}$	0.58	0.60	1.07	1.24	1.43	2.64
$\bar{p}$	.337	.249	.083	.095	.143	.092
Family Type 4. Average Income: \$1,941 per Year						
a	320	170	—18	—32	—65	—376
b	.192	.177	.078	.113	.166	.275
$\bar{\eta}$	0.54	0.67	1.13	1.17	1.25	3.38
$\bar{p}$	.357	.265	.069	.096	.133	.081
Family Type 5. Average Income: \$2,421 per Year						
a	400	260	—30	—70	—35	—525
b	.204	.127	.088	.133	.125	.323
$\bar{\eta}$	0.55	0.54	1.16	1.28	1.13	3.04
$\bar{p}$	.369	.234	.076	.104	.111	.106
Family Type 6. Average Income: \$2,249 per Year						
a	270	310	18	—10	—50	—540
b	.181	.112	.068	.093	.150	.396
$\bar{\eta}$	0.60	0.45	0.89	1.05	1.17	2.54
$\bar{p}$	.301	.250	.076	.089	.128	.156

V. COLUMBUS (OHIO): CLERICAL FAMILIES

EXPENDITURE ON:

	Food	Housing, Fuel, and Light	Recreation and Personal Care	Clothing	Transportation and Furnishings	Miscellaneous and Savings
Family Type 1. Average Income: $1,887 per Year						
a	195	195	—13	—40	—90	—245
b	.148	.150	.087	.111	.235	.269
$\bar{\eta}$	0.59	0.59	1.09	1.23	1.26	1.94
$\bar{p}$	.250	.253	.080	.090	.187	.139
Family Type 2. Average Income: $1,830 per Year						
a	265	220	11	—31	—70	—395
b	.143	.119	.079	.105	.224	.330
$\bar{\eta}$	0.50	0.50	0.93	1.19	1.20	2.92
$\bar{p}$	.288	.240	.085	.088	.186	.113
Family Type 3. Average Income: $1,943 per Year						
a	200	210	—35	—36	—100	—245
b	.185	.132	.099	.112	.222	.250
$\bar{\eta}$	0.64	0.55	1.22	1.20	1.31	1.97
$\bar{p}$	.288	.241	.081	.093	.170	.127
Family Type 4. Average Income: $1,931 per Year						
a	220	165	7	—9	—95	—290
b	.205	.145	.075	.098	.193	.284
$\bar{\eta}$	0.64	0.62	0.95	1.05	1.34	2.10
$\bar{p}$	.318	.232	.079	.093	.144	.135
Family Type 5. Average Income: $2,061 per Year						
a	235	200	29	—60	—140	—265
b	.238	.124	.063	.145	.190	.240
$\bar{\eta}$	0.67	0.56	0.82	1.26	1.57	2.16
$\bar{p}$	.353	.222	.077	.115	.121	.111
Family Type 6. Average Income: $2,170 per Year						
a	230	260	4	—13	—75	—405
b	.166	.122	.067	.094	.180	.371
$\bar{\eta}$	0.61	0.51	0.97	1.07	1.23	2.03
$\bar{p}$	.273	.241	.069	.088	.146	.183

DEMAND FUNCTIONS FOR MERCHANDISE AT RETAIL

ROSWELL H. WHITMAN
R. H. Macy and Company

I. INTRODUCTION

THE aim of this paper is necessarily a modest one. Only after years of effort on the part of many economists and statisticians and more than ten years of research by an extensive staff, was the late Professor Henry Schultz, to whom this book is a tribute, able to produce the monumental *The Theory and Measurement of Demand.* It is not likely that an extension of demand analysis into an entirely different field with marked limitations on the time and the personnel available for such study will yield definitive results. It is hoped, however, that the presentation of results, even though tentative and incomplete, will serve to stimulate analysis and research in a new and important field.

A study of demand functions for merchandise at retail might be concerned with the total takings of a commodity by consumers in a particular market or economy at different relative levels of retail prices for the commodity. This would represent a logical extension of the demand analysis already successfully applied to agricultural commodities. Such a study, however, probably awaits either more comprehensive family-budget data or a different type of market statistics. Another possible approach to the problem which is used in the present paper is to study the reaction of consumer takings to price changes at one retail outlet only. This approach abandons the assumptions of a perfect market and perfect competition which are so important in working out the classical theory of demand. It does not study the effect of price on total commodity flow in a "perfect" market but rather the effect of price change on competition in an imperfect market. In view of the importance of knowing more about the behavior of competition, however, the shift in emphasis is not necessarily to be regretted.

II. THE RETAILER'S POINT OF VIEW

The entire field of retail price behavior is very broad. In some cases prices are pegged by legal or other action and rarely change;

in others, the life of a particular piece of merchandise is so short that price moves downward only to clear excess stock. In the latter situation acceptance of the merchandise changes so rapidly that a measurement of the general reaction of takings to price would be impossible, because there is not that "routine of consumer behavior"[1] which is necessary for the measurement of any demand function. An analysis of the price behavior for staple merchandise, however, may prove more fruitful. The selection of staple merchandise for study and the limitation of the period analyzed to two or three months largely eliminates the possibility that changing tastes or consumer incomes will affect the demand relation. Thus, it is possible to avoid one of the most difficult problems in the measurement of demand. There will, of course, still be fluctuations in demand due to such factors as the weather. Such causes of sales variation, however, because of their more random behavior, are much less likely to obscure a price-sales relationship than changes in tastes or income.

Price fluctuation for fairly staple merchandise usually arises because one or more important retailers believe that there will be an increase in sales if the price is lower. A merchant seldom, if ever, questions the existence of a demand function with a negative slope. For some types of merchandise he may feel the slope is steep—an inelastic demand—and consequently price fluctuations will be relatively rare (except to clear stock as already noted). But in other situations his belief in a greater elasticity will result in frequent price manipulations.

If this is true, the price of a commodity in a single retail outlet will change from time to time either to meet the price of a competitor or because of a desire to experiment with a different price quotation. At each price there will be a varying volume of sales. The problem is to determine whether there is any uniform relation between the price and the quantity sold.

A priori it might be expected that no uniformity would exist unless all the other factors—prices of competing commodities, prices of competitors, etc.—which will affect the sales of a commodity could be controlled. But if the retailer is to act at all, and he frequently does experiment with price, he must assume that the factors affecting the sales of a commodity, in addition to the price, either will be of little importance quantitatively or will maintain some fixed relationship to the price of the merchandise in question. This attitude might be

[1] Henry Schultz, *The Theory and Measurement of Demand* (Chicago: University of Chicago Press, 1938), pp. 133–35.

called a pragmatic *ceteris paribus*. An actual search for price-sales uniformities may indicate the extent to which such an assumption is justified.

III. DERIVATION OF SIMPLE PRICE SALES RELATIONSHIPS

a) *The data.*—The data for each of the commodities on the basis of which price-sales relationships will be derived comprise the daily sales of each commodity together with the corresponding prices. The period covered is 2.5 months. The commodities included are staples—carried in stock year after year—and, as far as can be determined, without violent seasonal changes in demand. These commodities are standard merchandise which can be duplicated in a number of retail

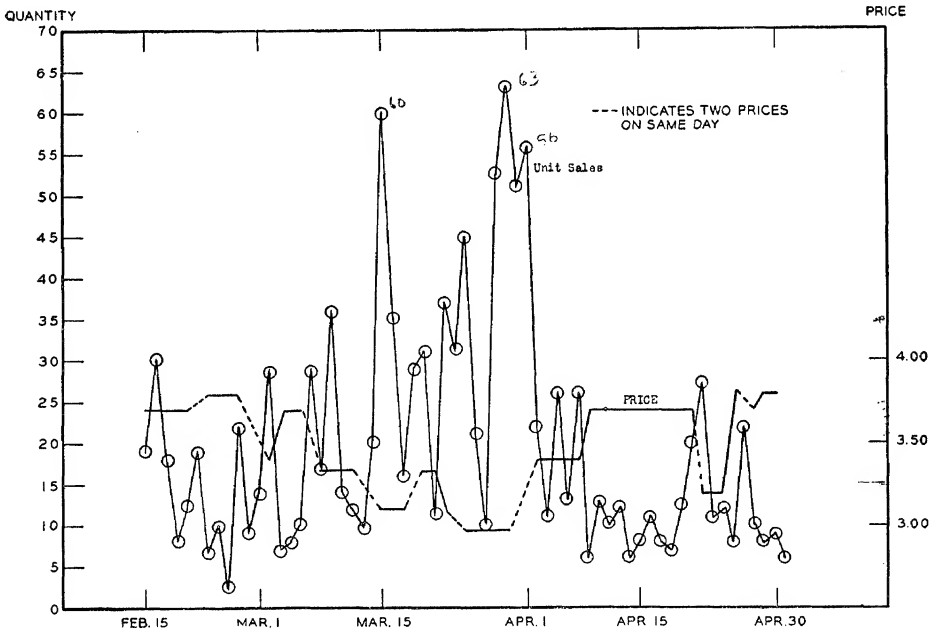


FIG. 1.—Daily unit sales and prices—Article A

stores. Partly because of this fact, price changes are more frequent than would be typical of less staple department-store merchandise.

The sales data will be subject to considerable variation independent of prices of competing commodities, etc., due to purely random causes as well as to the differing importance of various days of the week in retail selling.

Rather than make minute adjustments for such variation, it seems preferable to average the daily sales for a period of constant

price. The price-sales relationship will be determined for both the adjusted and the unadjusted data.

Figure 1 is a graphic representation of the type of data used. It shows the daily sales and price of a particular article of merchandise, hereafter known as Commodity A. The price changed fifteen times during the period. There is no marked trend in either the price or the sales series. Inspection of the chart indicates an inverse relationship between price and sales.

b) *The equations.*—The selection of the equations to express the relationship between sales and price is simplified by the fact that quantity taken is clearly the dependent variable. A price results from certain decisions by the retailer. The customers purchase a certain amount at that price. The supply is, at least under the conditions studied, perfectly elastic at each price set. This short-cuts the problems of shifting supply curves which has plagued statistical demand study.

The simplest assumption is that

$$q = f(p) \, , \tag{1.0}$$

where q is the unit sales each day and p the price on that day. Or, using adjusted data, we have

$$Q = f(p) \, , \tag{1.1}$$

where Q is the daily sales average for each period during which price does not vary.

Since we are not concerned at this point with the influence of other prices, etc., on sales, the only other variable which might be introduced is time. For a period of 2.5 months, time is a seasonal rather than a trend factor. As noted above, we have selected cases where the seasonal variation is relatively unimportant and not likely to affect sales elasticity to any marked degree. To test the possible effect of such a factor, however, we will experiment with the assumption

$$Q = f(p \, , t) \, . \tag{1.2}$$

For the specific form of the price-sales equation, only constant elasticity equations were used. The principle reason for this choice is the fact that such equations fit the data very well.

A straight line

$$q = a + bp \, , \tag{2.0}$$

invariably gives a poor fit to the data, while the equation

$$q = Ap^{a} \tag{2.1}$$

invariably gives a good fit.

The logic of an assumption of constant elasticity has been disputed by many observers. It has been claimed that, above a certain price, the demand for a firm's merchandise will be very elastic; below a certain price, quite inelastic. This reasoning is based upon the anticipated actions of competitors—their failure to follow a price rise and their duplication of a price cut. There has also been considerable discussion in retail circles of the theory of a psychological price —that is, a price at which maximum sales will develop and below or above which volume would be smaller. If this were true, the price-sales equation would certainly be more complex than is assumed in a constant elasticity equation.

Actual studies of price-sales relationships illustrated below do not support either argument. Possibly, however, neither argument should be applied to these particular relationships. Those who claim an increase in elasticity above a certain price may think of potential prices higher than any actually existing in the market. The theory of a "psychological" price may be most applicable to the price level of one commodity compared with the prices of competing commodities. The price behavior analyzed here, however, is that of a single item of merchandise which already has customer acceptance and for which the customary price is known.

Since, then, there is no well-established a priori reason why curves of constant elasticity may not represent the data under consideration, the criteria of simplicity, ease of computation, and goodness of fit suggest an experiment with the following constant elasticity equations:

$$q = Ap^{a}\,, \tag{2.1}$$

$$Q = Ap^{a}\,, \tag{2.2}$$

$$Q = Ap^{a}\, e^{\beta t}\,. \tag{2.3}$$

It will be noted that equation (2.3) states that, when price is fixed, the relative increment of quantity purchased per unit of time (day) is independent of the level at which price is fixed. The assumption is also made that this relative increment is constant during the period studied and that no more complex function of (t) is necessary to express the seasonal. This latter assumption does not have any a priori justification but has seemed adequate to describe the behavior of the data.

Another method of studying price-sales relationships is that of link relatives. This method is useful in obtaining an over-all elasticity for a group of articles similar in nature. In many situations the price of a single item will not vary often enough to make it possible to derive an equation in the form of (2.1), (2.2), or (2.3), unless the analysis is extended over a longer period with the resultant introduction of the effect of changes in tastes, changes in the merchandise, etc. This greatly restricts the possibilities of analysis unless it is feasible to determine the elasticity for a "commodity" comprised of a number of similar articles.

From the point of view of the merchant, such an approach is frequently the most useful. The price-sales behavior of a particular item which may be nonexistent in six months is of less importance than the price-sales behavior of a commodity group. For example, suppose that ten different brands of low-priced toilet soaps in twenty different sizes are being sold. The elasticity of sales for Brand C, which may not be carried another year, is less important to the retailer than the elasticity of sales for low-priced toilet soaps as a whole. The link-relative method makes it possible to generalize the price-sales behavior for all the ten brands and twenty sizes. The function will be

$$Q' = f(p') \,, \tag{3.0}$$

where

$$Q' = \frac{Q_i}{Q_{(i-1)}} \,, \qquad \text{and} \qquad p' = \frac{p_i}{p_{(i-1)}} \,.$$

Adjusted data are used for fitting this function, Q_i and p_i being the average daily sales and price for any period of constant price, and $Q_{(i-1)}$ and $p_{(i-1)}$ being the average daily sales and price for the period immediately preceding. The equation can be fitted to all the Q', p' observations irrespective of the particular brand or size involved.

The legitimacy of this procedure depends on whether sales elasticity is homogeneous for the articles included in a commodity group. This can be tested empirically by graphic analyses or, in certain circumstances, by measuring elasticity for the individual articles. In the analysis for agricultural commodities the heterogeneity of the commodity studied is frequently masked in the analysis. Cotton, wheat, etc., have many grades and prices for each grade. Obtaining an elasticity for cotton or wheat as such is no different from obtaining an elasticity for low-priced toilet soaps as such, except that in the latter

case the data permit a test of the reasonableness of a demand function for a commodity which is not completely homogeneous. The usefulness of obtaining an elasticity for a commodity group justifies experimentations.

For the commodities analyzed below, the function

$$Q' = f(p') \tag{3.0}$$

will be fitted in the form

$$Q' = Ap'^{a}. \tag{3.1}$$

The constant elasticity form is again primarily called for because of a consideration of goodness of fit.

Before the link-relative method is applied to a commodity group, it will be tested along with equations (2.1), (2.2), and (2.3) for a single article.

c) Statistical results for a single article.—The daily sales and price for Article A are depicted in Figure 1. There appeared to be a negative correlation between sales and price. This is more clearly depicted in Figure 2, which shows the average daily sales of Article A

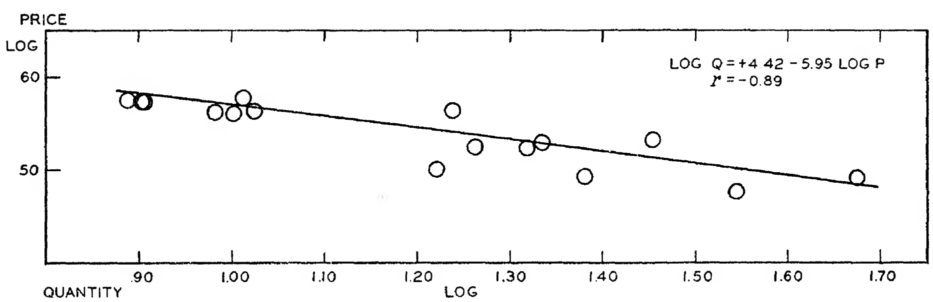


FIG. 2.—Relationship of average daily unit sales to prices—Article A

plotted against corresponding prices. Using equation (2.2) as the form of the price-sales function, the price-sales relation is

$$\log Q = 4.41 - 5.95 \log P$$

or

$$Q = 25710p^{-5.95}.$$

It will be noted that the sales elasticity is high.

Table 1 shows the results obtained in analyzing the price-sales behavior of Article *A* by fitting the equations (2.1), (2.2), (2.3), and (3.1) described above. The following points seem worthy of note.

TABLE 1

COMMODITY A: CHARACTERISTICS OF PRICE-SALES RELATIONSHIPS WITH QUANTITY AS DEPENDENT VARIABLE FOR A 2.5 MONTH PERIOD

q = Daily sales
Q = Rate of daily sale for period of constant price
p = Price
t = Time in days (origin at beginning of period)
Q' = Link relatives of rate of daily sale for period of constant price
p' = Link relatives of price

(Figures in parentheses are standard errors)

Equation	EQUATIONS				Descriptive Constants		
	Constant Term	log p	log p'	t	Elasticity of Sales η	r	Coefficient of Multiple Correlation Corrected for No. of Parameters R'
1. log q	4.22	—5.64 (0.85)			—5.64	—0.69	
2. log Q ...	4.41	—5.95 (0.78)			—5.95	—0.89	
3. log Q ...	4.47	—5.83 (0.66)		—0.0039 (0.0014)	—5.83		0.92
4. log Q'	—0.019		—6.64 (0.72)		—6.64	—0.92	

1. When adjusted data are used (i.e., average daily sales for each period of constant price), the coefficients of correlation are about 0.9, indicating a fairly good fit.

2. The third equation, (2.3), with time introduced as an independent variable, improves the fit only slightly.

3. The estimates of sales elasticity obtained by use of the first three equations are similar. The range is from −5.64 to −5.95, much less than the smallest standard error. The link-relative method in the fourth equation, (3.1), gives a somewhat higher elasticity, but in view of the size of the standard errors the difference may not be significant.

4. There is no question but that the response of sales to price is an elastic one. Any of the equations shows that a 1 per cent change in price results in a change in sales of over 5.5 per cent in the opposite direction.

d) *Statistical results for a number of similar articles.*—An additional test of the usefulness of these assumptions and methods can be obtained by extending them to a number of similar articles for the same time period. Table 2 shows the resulting elasticities when the four equations (2.1), (2.2), (2.3), and (3.1) are fitted to the data

for seven articles similar in character which comprise a commodity group. The fits are good except in equation (2.1), which uses the unadjusted data. For equations (2.2), (2.3), and (3.1) the coefficients of correlation (simple or multiple) range from 0.61 to 0.98, with the

TABLE 2

ELASTICITIES OF SALES FOR SEVEN COMMODITIES STUDIED, ACCORDING TO THE TYPE DEMAND EQUATION ASSUMED

q = Daily sales
Q = Rate of daily sale for period of constant price
p = Price
t = Time in days (origin beginning of period)
Q' = Link relatives of daily sale for period of constant price
p' = Link relatives of price

Equation 1 = $\log q = a + b \log p$
Equation 2 = $\log Q = a + b \log p$
Equation 3 = $\log Q = a + b \log p + ct$
Equation 4 = $\log Q' = a + b \log p'$

COMMODITY	TYPE OF DEMAND EQUATION ASSUMED			
	I	II	III	IV
A	-5.64 ± 0.85	-5.95 ± 0.78	-5.83 ± 0.66	-6.64 ± 0.72
B	-2.43 ± 0.85	-4.77 ± 0.47	-5.51 ± 0.18	-5.16 ± 0.41
C	-4.94 ± 1.21	-4.84 ± 1.62	-4.80 ± 1.71	-6.13 ± 1.49
D	-5.37 ± 1.33	-6.60 ± 1.92	-7.02 ± 1.44	-6.00 ± 1.30
E	-1.94 ± 0.81	-5.69 ± 0.69	-5.84 ± 0.67	-5.60 ± 0.39
F	-5.46 ± 1.20	-4.77 ± 1.84	-3.95 ± 0.32	-4.00 ± 1.06
G	-5.75 ± 0.66	-3.52 ± 0.44	-3.49 ± 0.59	-3.70 ± 0.62

majority being about 0.90. The comparisons of elasticities for the seven articles as calculated from the four equations show the following:

1. Except for the unadjusted data in the case of Article B, the differences in elasticities obtained for a particular article by any of the four methods are small when compared with the standard errors.

2. Almost all elasticities are over 4, indicating a marked reaction of sales to price change.

3. Except for Articles F and G, the difference in elasticities between the various articles are small when compared with the standard errors. This indicates a similar reaction of sales to price change for the various components of the commodity group.

4. Articles F and G apparently have somewhat lower elasticities than the other five. This is probably due to the fact that, while the two articles have a use identical to the other five, their customer acceptance is not so wide. This means that a price reduction will attract fewer new customers than in the case of more widely known articles.

e) *Stability of price-sales relationships.*—Although it is evident

from the above analysis that price-sales relationships can be measured, such measurements are not likely to be of much value unless the relationships are stable. Only if this is true can action be taken on the basis of past experience. It is important to know, therefore, whether sales reactions are sufficiently stable to make the results obtained from observations at one time period applicable to a later period. The results of one test of the stability of price-sales relationships are shown in Table 3.

TABLE 3

ELASTICITIES OF SALES FOR FOUR COMMODITY GROUPS IN TWO SUCCESSIVE YEARS FOR IDENTICAL TIME PERIODS (WITH STANDARD ERRORS)

Equation assumed

$$\log Q' = a + b \log p'$$

where

Q' = Link relatives of rates of daily sale for periods of constant price
p' = Link relatives of price

Commodity Group	First Year	Second Year
I	2.47 ± 0.61	3.21 ± 0.93
II	4.31 ± 0.94	4.50 ± 0.97
III................................	4.69 ± 0.58	4.32 ± 1.05
IV................................	5.34 ± 1.21	4.27 ± 1.44

The important points to be noted in this table are two. (1) Commodity Group I has a smaller elasticity than the other three groups, the reason for this probably being that this group was "out of season" at the time the analysis was made. Such narrowing of the market might be expected to lessen the reaction to price change. (2) From year to year the changes in the measure obtained for elasticities seem to be well within the limits of sampling error. Especially striking is the fact that the relation of the elasticity of commodity Group I and that of the other groups is maintained from one year to the next. This indicates considerable stability in the reaction of consumers to price changes.

While this one test does not prove that there is generally no change in consumer reaction to price change from year to year, it does indicate that sales elasticity for a staple commodity may well be relatively stable. General observation, in fact, indicates more stability than could be proved even by a much more extensive statistical analysis. Merchants do act on the basis of their experience as to the effect of short-run price changes. While errors do occur, the results are usually close to expectations.

IV. DERIVATIONS OF DEMAND FUNCTIONS HOLDING OTHER FACTORS CONSTANT

Price-sales relationships as reliable as those illustrated above are of obvious pragmatic value. From the practical point of view, also, they are a neat solution since they sum up the actions of customers to an entire market situation in terms of one variable. Many things change together with the change in price of one commodity—the prices of competing commodities, prices of the same commodity in competitors' stores, etc. Yet it is possible to measure customer reaction in terms of the one variable that can be independently manipulated. A straightforward basis for action is thus provided. If a price change were contemplated which was quite different from past experience, then explicit consideration would have to be given to what competitors would do. For the commodities here discussed this is not necessary.

It is obvious that these price-sales relationships are neither the market demand curves of neoclassical theory nor are they the demand curves which are assumed in the modern theory of differentiation and imperfect competition. This latter theory assumes a demand function with a slope significantly greater than zero when the prices of competing goods are held constant. In order to approach this type of demand function, it would be necessary to take explicit cognizance of the behavior of the prices of competing goods. In the case of the retailer this means the prices of this same merchandise in his competitors' stores as well as the prices of competing goods in his own establishment.

My experiments with this type of approach are not extensive enough to make certain what type of function would result if the process of keeping things constant were carried to its logical conclusion. They indicate, however, the possible results of such a process. For example, it is possible to measure the effect of introducing the prices of competing articles in the same store into the demand equations discussed above.

For the price of competing items it is necessary to use a price index, since in the situation analyzed a large number of articles have about the same position in the market. This makes it impossible to introduce each item into the regression equation. The movements of the index, however, represent fairly accurately the relative prices of competing articles.

The form of the equation used to test the effect of other prices is

$$q_A = Cp_A{}^{\alpha}\, p_I{}^{\beta}\,, \tag{4.0}$$

where q_A is the daily sales of Article A (unadjusted data), p_A the price of Article A, p_I an index of the prices of competing commodities, and C a constant. The introduction of time in the estimating equation does not improve the fit.

TABLE 4
ELASTICITIES OF SALES AGAINST PRICE AND AN INDEX OF PRICES OF COMPETING COMMODITIES FOR SIX ARTICLES: PERIOD 2.5 MONTHS

Equation assumed

$$\text{Log } q_A = c + b \log p_A + c \log p_I$$

where

q_A = Daily sales of Article A
p_A = Price of Article A
p_I = Index of prices of competing commodity
r = Coefficient of correlation
η = Elasticity of sales
(Figures in parentheses are standard errors)

Article	Elasticities of Sales		Coefficient of Multiple Correlation Corrected for No. of Parameters	Results with Price of Same Article as Only Independent Variable	
	Against Price of Same Article	Against Index of Prices of Competing Commodities		r	η
A	—5.94 (1.00)	+1.82 (2.96)	0.68	—0.69	—5.64 (0.85)
B	—10.5 (2.61)	+16.2 (8.69)	0.64	0.36	—2.43 (0.85)
C	—7.08 (1.35)	+7.79 (2.60)	0.58	0.50	—4.94 (1.21)
D	—9.03 (1.23)	+8.25 (2.87)	0.64	0.49	—5.37 (1.33)
F	—7.22 (2.15)	+7.32 (7.54)	0.51	0.52	—5.46 (1.20)
G	—6.47 (0.95)	+1.39 (1.29)	0.74	—0.74	—5.75 (0.66)

When this equation (4.0) is tested on the data for six of the seven articles already analyzed, interesting results are obtained, the most important of which are summarized in Table 4. This table shows the correlation coefficients and sales elasticity when the price of the article is the only independent variable, as well as the results obtained when the price of competing commodities is introduced.

The important points to be noted in this table are:

1. The simple coefficients of correlation between price and sales are not high, ranging from -0.36 to -0.74. (The high coefficients were obtained by use of the adjusted data.) It does not seem likely, therefore, that the introduction of an additional independent variable is necessarily meaningless due to possible correlation of random error.

2. The introduction of the Index of Prices of competing commodities improves the fit in three of the six cases. This indicates, therefore, that other factors—probably random in nature—are still very important in determining the sale of a particular article on a particular day.
3. In all cases the coefficient of regression of sales against the Index of Prices of competing commodities is positive, and in three cases very probably significant. This sustains the a priori expectations that high prices of competing commodities increase the sale of the alternative commodity and vice versa.
4. In all cases, the coefficient of net regression of sales against the price of the same article is greater after the introduction of the Indexes of Prices of competing commodities than when the price of the same article is the only independent variable. In three of the six cases the difference is greater than either standard error. This also confirms the a priori expectation that holding other things constant would increase the effect of price change on the sales of the item with the changed price.

While considerably more research would be necessary to establish the point completely, this experimental analysis seems to show that consumers are interested in price relationships between competing commodities and very likely will shift part of their patronage from items whose price is unchanged to those offered at a bargain. The tendency for sales elasticity to increase with the introduction of an additional price variable is also important. If other factors, particularly that of the price of the same item at competing retail units, were added to the equation, the effect of independent price change at one store on the sales at the same store might be still greater.

Whether the refinement of the statistical methods would finally yield a "demand function" with no slope is a matter for conjecture. It is undoubtedly true, however, that results would differ greatly according to the character of the merchandise studied. The whole question of the total effect of the demand patterns for merchandise at retail on the imperfections of the retail market requires further investigation.

V. CONCLUSIONS

Before summing up the possible significance of the results for economists, it is important to point out that their practical usefulness has not as yet been very great. While it is possible to establish a relationship between price and sales that is both precise enough to measure statistically and fairly stable, it has not proved worth while to try to do so on a large scale. The analysis is necessarily costly and time-consuming. Usually the retailer is interested only in whether a proposed price manipulation "will pay." An answer to this can frequently be found without elaborate study.

It is still worth while to discuss the theoretical implications of the results. First, it might be noted that the most precise and fruitful relationships are established when no attempt is made to "keep other things constant." While the introduction of other variables probably increases the elasticity of sales of an item against its own price, the resulting function is at the best of theoretical importance.

But what meaning has a price-sales relationship when the prices of some competitors change and those of others do not, when the prices of some competing commodities change and others do not? Yet this will be of frequent occurrence in the imperfect retail market. The imperfection of this market results from a complex mixture of differentiation and oligopoly. The fewness of sellers, that is, the fewness of big sellers, means that competitors' reactions will be taken into consideration in the setting of price, while the ever present differentiation means that retailers have alternatives in their reaction to price changes by their competitors. They may or may not duplicate a price change, and in practice some do and some do not.

In these circumstances, if an important group of competitors finds that a commodity has a high sales elasticity, it would indicate to the economist that there probably will be a vigorous price competition. An elasticity of 5 makes price a tempting weapon in the search for maximum profits. Of course, if the elasticity becomes very large, it makes out-and-out price competition too dangerous to use; but this is generally unlikely in the present retail market. A small elasticity leads to the use of other competitive weapons, such as selling cost, in the search for maximum profit. As economists explore the problem further they will probably find that where merchandise has a high sales elasticity, selling cost is relatively small, and vice versa.

It would seem probable, therefore, that price-sales relationships of the type analyzed above are of importance in the general problem of equilibrium and allocation of resources in retailing. The "all other things being equal" demand functions for a single product and firm are a nicer theoretical concept, but they are difficult to apply where, as in retailing, the type of "imperfection" cannot be neatly classified. In such circumstances the more empirical definition of "demand" suggested here may be of theoretical as well as of practical importance.

DEPARTMENT-STORE COST FUNCTIONS[1]

JOEL DEAN
University of Chicago and the Cowles Commission

I. INTRODUCTION

AN EMPIRICAL study of the operating behavior of the individual enterprise should supply factual information for verification of theoretical formulations. This is a counsel of perfection, however, and at the present level of development of economic science much of the value of this type of empirical study lies (1) in determining whether the assumptions underlying the economic theory of the individual firm are sufficiently typical, extensive, and realistic to have descriptive usefulness and to admit of empirical verification and (2) in developing statistical techniques which will in time permit the effective use of the results of theoretical research in the solution of economic problems.

The systematic theoretical analysis of the form of the cost function of the individual firm was a by-product of the attempt to clarify the relation between cost curves and industry supply curves under the assumption of perfect competition. This assumption, however, was not only unrealistic in an institutional sense, but it oversimplified the problem of cost and output determination by implying that the only part of the marginal cost curve relevant for output decisions is the rising section. The development of the economics of imperfect competition represented, therefore, a significant improvement in the theoretical background of empirical research, since by dispensing with the postulates of perfect competition it reduced the assumptions necessary to the theory of costs to the familiar technical laws of diminishing returns and variable proportions, and the psychological postulate of economically rational behavior.

[1] I wish to express my indebtedness to Miss Phyllis van Dyk for important assistance in conducting the study and preparing the paper, and to Miss Mary Hilton Wise for assistance in revising the manuscript for publication. Professor Arthur F. Burns, Mr. R. W. James, Mr. John Smith, Professor Gerhard Tintner, and Professor Theodore O. Yntema were kind enough to read the manuscript and make valuable suggestions. The Social Science Research Committee of the University of Chicago and the National Bureau of Economic Research provided funds for the conduct of the study. I am particularly grateful to Mr. Oswald W. Knauth, president of the Associated Dry Goods Corporation, and to the officials of the Lord & Taylor department store for their generosity in making these data available for analysis.

When the assumption of perfect competition was dispensed with, however, it immediately became necessary to analyze that category of expenditure which was devoted to influencing the individual firm's demand function. The conditions of equilibrium for the individual firm were restated to include selling cost as a variable category of expenditure exhibiting a functional relation to sales.[2] In the formal solution of the equilibrium condition, it is assumed that selling expense and expenses of production in the narrow sense can be separated and the amount of each determined independently. By definition, the two may be clearly distinguished on a functional basis; selling costs are accordingly expenditures designed to alter the shape or position of a firm's demand function, while costs of production proper consist only of those outlays technically necessary to produce a given physical quantity of a particular product.[3]

Empirically, it may in some cases be possible to perform the operaations which are required to separate cost data into these two categories. When the business organization whose costs are being analyzed is specialized to supplying selling service, however, the basis of the distinction itself becomes uncertain. This is true because the assumption made in theoretical writings that a "given demand" for a good exists apart from the conditions under which the good is presented to the public is only realistic for very homogeneous and frequently purchased commodities. More usually, it is impossible to isolate "demand" from selling or promotional cost which create or modify it.

Promotional activities designed to crystallize and focus demand constitute a large part of the total operations of a department store. To some extent such promotion may quite properly be regarded as a utility-creating function, constituting an integral part of output of a retail store, namely, its sales service.[4] Moreover, promotional activ-

[2] See H. Smith, "Advertising Costs and Equilibrium," *Review of Economic Studies*, Vol. II (1934–35), and R. M. Shone, "Selling Costs," in the same volume. On polyperiodic production see Sune Carlson, "A Contribution to the Pure Theory of Production" (Ph.D. dissertation, University of Chicago, 1936), chap. vi.

[3] Shone has used the term "allocated costs" for costs of production because they "vary directly with output, and their redistribution, increase, or decrease for any given output is excluded by hypothesis. On the other hand, the unallocated expenditures (selling costs) may be incurred in any form, increased or decreased, in order to alter the price and the output of the product defined by the allocated costs" (*op. cit.*, p. 225).

[4] Particularly for goods that are purchased infrequently, the advertising, displays, demonstrations, and advice of salespersons required before a vague desire is transformed into a sale of goods may be viewed as creating a utility for which consumers are willing to pay. Even newspaper advertising may be considered as contributing to the store's sales service since it acquaints the consumer with prices and qualities of goods available, and since its "tone" may augment the prestige satisfaction derived from the merchandise.

ities pervade every accounting category of department-store cost and contribute to every aspect of stores' sales-service product. Thus even if the basis of distinction were clear, the practical difficulties of extricating promotional costs from pure production cost in the data would probably be insuperable.

The solution adopted for this difficulty was to consider all merchandising costs as costs of production. A more general argument may be made out for this procedure on the ground that selling expenditure can only augment the sales of the affected commodity by increasing its utility to consumers; more strictly, selling expenditure converts a given physical commodity into a different and more desirable economic good, and may accordingly be regarded as a cost of production in its most rigorous sense.[5]

Whatever may be the merits of this argument, its acceptance points to an additional source of embarrassment and difficulty in this study. If we are to consider that selling expenditures change the character and utility of the product, the unit of output is no longer homogeneous as selling costs vary. A cost function derived from accounting data over a period of time during which selling expenditure constituted a varying proportion of cost would therefore relate to a product which was not homogeneous throughout the sampling period.

This nonhomogeneity of output arises from the fact that promotional cost does not bear the simple technical relation to output that pure production cost does. The amount of selling costs it is advantageous to incur presumably depends upon (1) the position and slope of the firm's demand function and (2) the effectiveness of selling expenditure in modifying this demand function, rather than upon technical conditions of production. In the present case, however, since selling costs and costs of production are inseparable, the quantities of output and of selling expenditure cannot be determined independently, although some variation in the proportion of selling expenditure to strictly productive expenditure is undoubtedly possible. Especially in a department store dependent upon its prestige, promotional expenditures for spacious salesrooms, sales advice, and elaborate display are an essential requirement for the maintenance of sales volume. The situation approximates closely joint production in industry.

[5] Cf. Frank H. Knight: "In so far as they [changes in wants] result from a deliberate expenditure of resources, they become as all other economic operations . . . . . In fact, as we have previously observed, the advertising, puffing or salesmanship necessary to create a demand for a commodity is causally indistinguishable from a utility inherent in the commodity itself" (*Risk, Uncertainty and Profit* [Boston, 1921; reprint, London, 1933], p. 339).

The possibility that the proportion of selling to production costs may be kept fairly stable by certain institutional conventions or standards should therefore not be overlooked.[6] In such a case, the output would be fairly homogeneous, and the cost function discovered would be more likely to possess the stability necessary for purposes of prediction.

To recapitulate, the ambiguity surrounding the concept "unit of sales service," which may properly be considered the product of a retail enterprise, made it impossible to separate selling costs from costs of production proper. The inclusion of a variable and perhaps indeterminate quantity of the former results in heterogeneity of output through time. The cost function determined by our procedures cannot therefore be compared precisely with the static model of cost theory. All that can be claimed is that it shows the relation between a composite cost and some admittedly imperfect index of output.[7] Its heuristic value will, therefore, be confined to providing a description of merchandising costs when certain operating conditions have been stabilized in order to isolate the influence of output on cost.

As this suggests, with the exception of the inclusion of selling costs among costs of production, every attempt was made to make this empirical situation investigated conform to the static model for the short run. The investigator must wrest from his dynamic data something as nearly approaching a static function as possible if he wishes to find cost or demand functions which are conformable to those upon which so much valuable theoretical elaboration has been based. Such a procedure increases the usefulness of the findings for purposes of estimating costs and also makes their interpretation more clear. Once the basic relation between costs and output has been established, it becomes relatively simple to adjust for the effect of changes in operating conditions, such as shifts in wage rates, material prices, etc., when desired. Of course, changes which lead to factor substitution cannot be handled in this way, since the cost function itself will be altered.

The main problems in determining the cost-output relation arose from the necessity of breaking down available accounting data into the categories of fixed and variable costs as defined in theoretical writings and the selection of procedures for "rectifying" the data in

[6] Considerable stability in the store's standards with respect to sales personnel, fixtures, delivery, returns, and internal displays would be expected in view of the importance of prestige in this particular store's merchandising policy. Prestige is an important aspect of product differentiation in the oligopolistic department-store market in which this firm competes.

[7] For a further discussion of the index of output see Sec. II, Part C.

order to remove the effect upon cost of changing conditions of production. No completely satisfactory solution could be found to these problems in some instances, even though no attempt was made to analyze the cost functions of the department store as a whole.[8] Only three carefully selected departments in a large metropolitan store were studied, and in these departments no analysis was made of the behavior of aggregate operating costs. Since the allocation of overhead costs to these units is necessarily arbitrary, it was obviously impossible to make any economically significant estimate of the department's true aggregate cost. Fortunately, the critical questions of cost theory center on marginal rather than average costs, although a knowledge of average costs is indispensable to the individual firm in computing its profits. In the present study, therefore, all allocated general store overhead costs were excluded. The analysis was confined to the following cost elements: (1) advertising, (2) salespeople's salaries, (3) other departmental salaries, (4) inside delivery, (5) outside delivery, and (6) direct departmental expense. Obviously some of these departmental costs are to a degree fixed, but the inclusion of such items does not affect the determination of marginal cost and yields average cost behavior useful for certain types of managerial decisions. The resulting estimates of marginal costs are subject to error because of the failure to include certain general store expenses which may bear some relation to the departments' volume. Moreover, some items which have been included may have little relation to volume but are administered, allocated, or recorded in such a fashion as to show a spurious correlation with sales volume. For example, several of the indirect advertising cost items which are allocated to departments on the basis of dollar sales may appear to possess a closer relation to volume than is in fact the case.[9] Similarly, the closeness of the relation of inside delivery cost to volume may be exaggerated by the method of computing this cost.

II. METHODOLOGY

The major methodological problems in deriving empirical cost

[8] Two papers concerning empirical studies of other department stores should be mentioned in this connection: "Cost Functions in Merchandising," delivered by Miss Phyllis van Dyk at the 1940 research conference of the Cowles Commission in Colorado Springs, Colorado, and "Cost Functions in the Department Store," delivered by Roswell H. Whitman at the 1940 annual meetings of the American Economic Association (Philadelphia).

[9] The indirect advertising costs allocated to departments on the basis of dollar sales include direct mail, compositions and cuts, art work, interior decoration, car advertising, Foster Bureau, fashion shows, magazines, addressograph, window-dressing and sign-writing, salaries of advertising office.

functions from accounting records arise from the necessity of purging the data of the effects of dynamic influences which were at work during the observation period.

In order to get an adequate number of observations, monthly data were analyzed for sixty consecutive months, covering the years 1931–35 inclusive. In attempting to isolate the static components of these data, three chief methods of freeing the cost data from the distorting influences of extraneous variables were employed. First, by careful selection of the establishment, departments, cost items, and time period investigated, certain factors which would have obscured the cost-volume relation were held constant. Second, the influence of another group of factors was allowed for by using them as independent variables in the multiple-correlation analysis. Third, the effects of several disturbing forces were removed directly from the data by familiar rectification procedures.

A. COLLECTION OF DATA

1. *Selection of establishment.*—Among distributive enterprises department stores are of particular significance because of their quantitative importance in distribution and because of the variety of merchandise carried. In searching for an establishment to co-operate in a study of this type, a large metropolitan department store was found which had kept comparable and unusually comprehensive accounting and statistical records over a period of years. Furthermore, the management was sufficiently interested in research to understand the importance of this study to economics and to see its practical usefulness in business administration.

2. *Selection of departments.*—Selection of the selling departments which would be most suitable for statistical cost analysis was made after an examination of each department with reference to the following criteria:

1. The sales volume should vary from month to month in such a fashion as to give a wide range of volume and a fairly uniform coverage of this range.
2. The heterogeneity of the merchandise sold in the department should be at a minimum, so as to simplify the problem of finding a suitable output index.
3. The character of the merchandise and the nature of the transaction should be relatively uniform from month to month, in order to minimize changes in the meaning of the output index through time.
4. The department should be relatively large, to maximize the managerial significance of the study and minimize the effects of indivisibility of input factors.
5. The changes in layout, general method of operation and managerial person-

nel during the period of analysis should be at a minimum, in order to approximate short-run cost functions by holding technology and plant scale as nearly constant as possible.

On the basis of these criteria, three departments were selected for study: hosiery, women's shoes, and medium-priced women's coats.

3. *Selection of the time period for analysis.*—Since the objective of this investigation is to determine statistically the behavior patterns of short-run selling cost, it is desirable to study the effects upon cost of changes in operating conditions in a situation which precludes the possibility of bringing certain of the input factors into optimum adjustment—a situation in which some cost elements offer effective resistance to adjustment to the prevailing operating conditions. Analyses of monthly or weekly observations of cost and operating conditions satisfy this condition, since many input factors in the department store cannot be adjusted to these relatively rapid changes in output and other circumstances. Monthly observations appeared more suitable than weekly observations for this purpose since fewer arbitrary cost allocations are involved because most of the statistical and accounting records of the firm were in terms of months. In addition, random and irrelevant fluctuations are more likely to be averaged out by taking the longer period.

In determining the years to be selected for study, the following criteria were applied:

1. Changes in the space occupied, in layout, in general methods, and in the management personnel should be at a minimum.
2. Accounting records should be comparable throughout the period of analysis.
3. A sufficient number of months should be included to permit valid statistical analyses.
4. There should be sufficient independent fluctuations in demand, and consequently in output, to give a fairly wide range of observations of both cost and output.

Using these criteria as a basis for selection, the best years for analysis appeared to be 1931, 1932, 1933, 1934, and 1935, which made available sixty monthly observations.

B. DIRECT RECTIFICATION OF DATA

Changes in operating conditions which could be removed directly are:

1. Time lag between the recording of cost and of the volume of output to which the cost contributed
2. Variation in the number of selling days in a month

3. Changes in salary rates and material prices
4. Seasonal variation not associated with sales volume

1. *Time lag.*—The time lag between the recording of cost and of operating conditions which give rise to the cost was found to be appreciable for advertising only. During the years in question, however, this lag had been foreseen by the store and to a certain extent removed from the data in the process of recording the cost. Any advertising which occurred on the last day of the month was charged to the following month, on the supposition that it was to the sales of the second month that such expense contributed. This procedure assumes that the largest part of the effect of advertisement is felt on the day following its appearance and that the part which is carried over to the subsequent day is so sharply diminished as to be negligible for practical purposes. Consultation with store executives indicated a consensus that this assumption is correct. The removal of a recording lag, therefore, appeared unnecessary.

2. *Number of selling days.*—Variation in the number of selling days per month undoubtedly exerts an influence on the magnitude of cost, but the relation is neither proportionate nor constant because of the differential effect of length of month on the several subdivisions of total variable cost. Salaries of buyers are constant regardless of the length of the month. Salaries of the salespersons, stock persons, and clerical help are paid on a weekly basis and allocated to each month on the basis of the proportion of the overlapping week falling in the month. These costs, however, are not directly proportional to number of selling days, because of holidays, for which full salaries are paid. Other items, such as inside delivery, outside delivery, and departmental direct expense, vary somewhat with number of selling days, but this variation results indirectly from the influence of the number of selling days on the volume of sales rather than directly from the length of the work month.[10]

It may be concluded, then, that the number of selling days is a factor affecting both the dependent variable, cost, and certain independent variables. But since none of these relations can be considered proportional, the number of selling days should, therefore, be employed as an independent variable in the multiple-correlation analysis, in order to avoid the distortion and spurious correlation which would arise from proportional correction in the original data.

[10] Sales volume is probably not proportional to the number of selling days, since the occurrence of holidays often results merely in a time redistribution within a relatively short period and not a loss of sales. Hence, even the cost items most closely related to volume do not vary proportionately with the length of the month.

3. *Changes in input prices.*—Changes in salary rates and material prices are also likely to cause variation in cost behavior over a period of time. In determining whether correction for these changes is necessary and, if so, how it should be effected, several important questions arise:

1. Were there significant changes in wage rates and material prices during the period under study?
2. Are the observed changes in rates separable from the changes in cost which are due to variation in volume of sales?
3. Can an index of wage changes be found that will reflect changes in the labor market alone and not the fluctuations associated with improvement of the skill of the individual?[11]

An analysis of the variations in the salaries of salespersons over the period under consideration revealed that the average salary in each of the departments declined throughout 1931 and 1932 and most of 1933, then remained fairly constant, with a slight tendency to rise during 1934 and 1935. The magnitude of this fluctuation, however, was not great. Furthermore, it was not a blanket reduction but was highly individualized. The desirability of removing the effects of this change in average salary from the cost data depends essentially upon whether the salary cost per unit of sales was significantly affected. If this cost has not changed during the period, then the adjustments in salary rates have been approximately proportional to the changes in sales volume and have thus brought about what is, in effect, a constant piece rate. A study of the average salesperson's salary cost per dollar of sales for each of the three departments showed that this rate did remain substantially constant. On the basis of this evidence of approximately constant factor prices per service unit, it was concluded that no correction was needed.[12]

Inside delivery cost figures were not corrected because a large proportion of their total was made up of salaries which were so ad-

[11] It is important to distinguish between changes arising from career advancement of the individual (i.e., improved selling effectiveness) and those associated with changes in factor prices.

[12] Rectification of other kinds of salaries was regarded as undesirable for several reasons. Variation in the salary of buyers and assistant buyers in the departments in question over the years studied was small. In two departments the same buyer, at almost the same salary, was employed during practically the entire period. In the other department the salary changes which accompanied the change of buyers were not great. The changes which did occur were supposedly a reflection of increases in the "effectiveness" of the buyer rather than changes in the labor market. However, the buyers' labor market is so imperfect and the measurement of "effectiveness" of merchandising executives so ineffective that little solace can be derived from this possibility. Nevertheless, the absence of objective bases for rectification of these costs made it seem more questionable to try to correct them than to leave them as recorded.

ministered as to defy index-number correction for the same reason discussed under salary rectification and because the materials used in this activity were so diverse that construction of an accurate price index would have been more costly than the improvement in accuracy would justify in view of the relatively small magnitude of these expenditures.

Correction of outside delivery, which was all handled by an independent delivery company during the period of study, was unnecessary because the package rate had remained unchanged over this period.

Advertising cost was found to vary mainly in response to changes in newspaper linage rates. Of advertising expenditure, 70 per cent was for newspaper advertising, and an average of 83 per cent of this proportion was paid for actual space (the remainder being for copywriting and illustration); so that about 60 per cent of advertising expense was proportionately dependent upon linage rates. The three departments in question were remarkably similar to the store as a whole with respect to the distribution of advertising expense.

Since there were significant changes in newspaper rates over the period of analysis, it was necessary to devise a correction index to eliminate the effect of these variations. A monthly index of the composite space-rate fluctuations for all the newspapers in which advertising was purchased was constructed and applied to 60 per cent of the original advertising cost data. These corrected data were then combined with the uncorrected 40 per cent to give a new series of cost figures from which the influence of rate changes had been removed.[13]

From the above discussion it is clear that rectification of most cost items for changes in wage rates or prices was found to be either unnecessary or unpractical so that only advertising cost was rectified for rate changes.

4. *Seasonal variation.*—Although seasonal variation in cost may be regarded as deriving primarily from fluctuation in the volume of sales, it is to a lesser degree attributable to changes in other conditions. Since in this study volume is the principal causal factor to be associated with cost, only the seasonal variation not arising from changes in physical volume and in average value of sale should be removed

[13] The index employed was a weighted arithmetic average of price relatives based on January, 1931, rates, with total linage purchased in the base year as the weighting factor. Although these weights are not constant throughout the five-year period for each newspaper, the composite result differed negligibly from an index weighted with the five-year average annual linage.

from the data. It was decided, therefore, first, to study the net relation of cost to number of transactions and average value of transaction, and then to isolate and remove that part of seasonal cost not correlated with this relation.

The familiar Bean-Ezekiel technique of establishing the net regression of cost on number of transactions and average value of transaction was employed.[14] Cost residuals from these net relations exhibited a seasonal pattern. In order to correct for this, an additive type seasonal index was computed,[15] which was used to correct the costs of the coat and shoe departments. The seasonal pattern of the hosiery department was not sufficiently clear to warrant such correction.

Because of the regularity of seasonal changes in many aspects of department-store merchandising, the seasonal index arrived at in this way appears to correct the cost data for a number of distorting influences which could not be successfully removed individually. Differences in the amount and character of sales service performed, for instance, appear to have for some departments a regular seasonal pattern which accords with the observed net seasonal variation in costs.[16]

Although variation in the number of selling days in a month was not a strong enough influence to show significant correlation with cost, its effect may be partly reflected in the residual seasonal pattern. February, a short month, shows higher than average costs in all departments. July and August, short months because the store closed Saturday, show high seasonal costs. Lags between the month in which sales effort was expended and the month in which the transaction is recorded appear to differ from month to month in a fairly regular pat-

[14] Cost residuals from these net relations were first plotted as a time series to ascertain whether there was any trend in the residuals or any perceptible shift in the seasonal pattern during the period of study, or whether the magnitude of the residual deviation was systematically different for different years. This analysis indicated that no significant trend, shifting seasonal, or variation in absolute magnitudes characterized the years studied.

[15] The index was constructed by taking the arithmetic mean of the absolute deviations for each month. This figure was then added to or subtracted from the original observations. The use of an additive, rather than a multiplicative, index was the logical outgrowth of employing an additive process in establishing the relationship of cost to number of transactions and average value of transaction. Moreover, the preliminary analysis indicated that the magnitude of the residual deviation was not proportional to the total cost, which further confirmed the choice of an absolute rather than a relative method of rectification.

[16] E.g., lower standards of service during the Christmas rush and during the January clearance sales may partially explain the low costs of these two months in the coat department. Differences in the character of the merchandise are also seasonally regular and would be expected to affect cost. The low costs of May and June in the coat department, for instance, are attributed by executives to sales of summer wraps which require relatively little sales service.

tern. For example, February, being the beginning of the style season for coats, carries missionary work and even actual sales completion for transactions not recorded until the following month. Similarly, November is loaded with some expense of training new girls or carrying over trained personnel preparatory to the holiday peak. Vacations, which caused cost distortion difficult to remove directly, appeared also to be included in this blanket rectifier, for they occur only in the months of July and August, and in these months costs are seasonally high. Thus the blanket correction of seasonal variation appears, happily, to cover a variety of specific causes for cost distortion which could not be removed individually.[17]

C. ANALYSIS OF RELATIONS

1. *Selection of index of output.*—In choosing the measurement unit that best represents output of retailing service, both physical volume of sales and dollar volume of sales were considered. The index that most accurately reflected the effect of physical volume upon the cost of the departments appeared to be the number of transactions. Number of units seemed less acceptable since sales-service effort was not believed to increase proportionately in selling several units in one transaction. Sales service, moreover, seems to be directly associated with the number of transactions when a department has a fairly homogeneous "product"; and some items of cost, such as inside and outside delivery, are so incurred and allocated as to be closely related to transaction volume.

An alternative measure of output, dollar value of sales, possessed some advantages because some items of cost, such as advertising, appeared more directly related to value of sales than to number of transactions. On the whole it was felt, however, that value of sales did not represent as acceptable a measure of output as number of transactions, especially as it was possible to reduce monthly figures for number of transactions to a comparable basis by holding average value of transaction constant. While a complete correlation analysis was made using value of sales as the output index, the results were not as satisfactory as those obtained by using number of transactions with average value of transaction held constant. These results will not be presented in detail here.[18]

[17] The effect of this correction as well as of most rectification is to increase the correlation between costs and sales volume.

[18] Cost deviations from the net regression curves of number of transactions and average value of transaction were plotted against dollar volume of sales in order to determine whether this factor was significantly correlated with the resi-

The size (money value) of the transaction had a clearly defined influence upon cost, even though its variation arose both from the number of units sold and from the size (money value) of these units. Records for the period of study did not permit a precise distinction between the two sources; but it seems evident that for the coat and shoe departments most of the fluctuations in cost came from differences in the value of units, whereas for the hosiery department most was attributable to differences in the number of units sold. Regardless of which of these influences predominates, the size of transaction probably affects cost materially. Even when the number of units is the same, more time is likely to be required for the sale of an expensive item than of a cheaper one.

The only available index of size of transaction was the average dollar value per sale. Since this measure reflects the effect of variation in dollar sales when number of transactions is held constant, it indirectly introduces value volume into the multiple-correlation analysis. This index is, however, subject to the same defects as dollar sales, namely, that it reflects changes in the retail price level, shifts in popularity among price lines, and changes in the nature of the article sold. Nevertheless, it appears to be a fairly satisfactory measure of an important cost influence.[19]

2. *Selection of additional independent variables.*—In addition to output as measured by volume of transactions, and average value of transaction, the following cost influences were tentatively selected as independent variables for the correlation analyses on the ground that they appeared to influence cost significantly and that they could be measured and treated statistically, but that their effect could not be safely removed by rectification of the original data.[20]

---

dual variation in cost not accounted for by the net regression curves. The application of this test indicated that the inclusion of dollar volume of sales as an additional independent variable in the multiple-correlation analysis was unnecessary. The influence of number of transactions and average value of transaction upon cost accounted for all the cost variation associated with dollar volume of sales, leaving no net correlation between the cost residuals and the latter.

[19] As has been noted, differences in the character of transaction were held to a minimum by selecting departments whose products were relatively homogeneous and whose sales service was similar for different types of articles. A large part of the remaining variation was probably removed by the use of a monthly time interval and by correction for seasonal pattern. Nevertheless, it is likely that important residual differences remained, part of which are reflected in average value of transaction.

[20] Two other cost influences were given consideration: rate of stock turnover and the proportion of inexperienced workers employed. Inquiry indicated, however, that the effect of inventory turnover was likely to be confined to cost and profit items not included in the analysis, e.g., interest, storage space, and markdowns.

1. Difference between actual and anticipated volume
2. Intramonth variability in volume of transactions
3. Number of selling days having unfavorable weather
4. Number of selling days per month
5. Change in volume from previous month
6. Fluctuations in business conditions

a) *Difference between actual and anticipated volume.*—The departure of actual sales from anticipated sales seems to have an important effect upon the costs of a department largely because of the difficulty of adjusting the working force to a violently fluctuating sales volume. The existence of a versatile contingent sales force which can be shifted among departments and of a trained alumni corps available for part-time work makes possible fairly accurate adjustment of salespersons and stock personnel to the sales volume to the extent that sales can be correctly forecasted. But if actual sales fall below planned sales, the adjustment is defective, and cost per sale rises.

The difference between budgeted sales and actual sales was considered as a possible measure of inaccurate forecasting but was subsequently found unacceptable because of incomplete records. The most promising of the alternative indexes of forecasting inaccuracies appeared to be the ratio of each month's sales to the sales of the corresponding month of the preceding year, since the earlier year's volume is usually a most important consideration in making plans for the current year's sales.

b) *Intramonth variability in volume transactions.*—Another important cause for variation in cost was thought to be fluctuation in sales volume within the time period selected as the unit of analysis. This fluctuation is of three main types: week-to-week variation, day-to-day variation within the week, and hour-to-hour fluctuation within the day. Fluctuation in volume affects cost partly because of the difficulty of adjusting the sales force to changes in need for its services. Intramonthly fluctuations may influence month-to-month differences in cost in two ways: (1) by departure from a predictable pattern of

The proportion of inexperienced salespeople employed did not appear to exert a serious influence on cost for several reasons: (1) the low rate of labor turnover in these departments; (2) the store's policy of developing a reserve of experienced extras and a versatile flying squadron of salespeople trained to work in several departments; and (3) the policy of starting inexperienced girls at low salaries and raising salaries only as their sales increase.

No suitable index of inexperienced personnel could be found. Average length of service is an inexact measure because of marked individual differences in learning speed and because it does not reflect a new clerk's previous training in selling the same merchandise at other stores. Selling cost per salesperson was equally unsuitable, since there appeared to be no correlation between length of service and the ratio of salary to sales.

variation or (2) by such extreme irregularity of variation as to make adjustments impossible even when correctly anticipated.[21] An examination of each type of fluctuation indicated that variability of sales within the month would probably fail to account for differences in cost between months.[22] Nevertheless, an index of variability was computed for one department to test this conclusion. Day-to-day fluctuation was chosen as most likely to be important and easiest to measure. An index of the variability in number of sales transactions was constructed by expressing the difference between the highest day and the lowest day of the week as a ratio of the lowest day.[23]

c) *Weather.*—Weather was thought to influence cost not only through its effect upon sales volume but also through its effect upon the predictability of sales. Bad weather, by suddenly driving sales below estimated volume, may make unnecessary the services of salespersons and stock persons previously employed.

Two kinds of adverse weather conditions can be distinguished: unseasonable weather and disagreeable weather. Since the former's disturbing effect upon cost was probably not great,[24] and since it was difficult to obtain an objective measure of unseasonable weather for the period under study, measurement of this phenomenon for use as an independent variable was abandoned.[25]

Disagreeable weather appears to affect costs by causing large day-to-day variations in sales which are not accurately predictable

[21] To a certain extent, successful adjustment depends upon ability to forecast sales accurately. Nevertheless, the problem is larger than this, for it also relates to the limited divisibility of production factors; e.g., the hour-to-hour fluctuations in business cannot be completely met by hourly adjustments in the selling force.

[22] Although the pattern of hour-to-hour variation had not been accurately determined, it was thought by executives to be approximately the same from month to month. The day-to-day variation within the week had been subjected to an analysis which showed that, although the fluctuations were marked for the departments under study, the pattern did not appear to differ significantly from month to month. Week-to-week fluctuation showed marked differences among months; but this type of variability could be more easily coped with because regularity of the seasonal pattern made it more exactly predictable and because a week is a long enough period to permit tolerably accurate adjustment of personnel needs.

[23] This method was based on the assumption that relative variation within the week was more significant than absolute variation. To have measured deviations from the expected pattern of fluctuation might have been preferable but was considered too laborious.

[24] Whether unseasonable weather results in actual loss of sales or merely in postponement is of no consequence here, for concern is only with the extent to which it may cause errors in planning personnel needs.

[25] Average variation of temperature from the normal for past years is not satisfactory, for it does not spot the timing and number of unseasonable days; while buyers' records are incomplete and are likely to rationalize unfavorable showings.

early enough to permit adjustments of selling and stock force. An index was constructed by tabulating for each month the number of days with extreme temperature (above 85° or below 15°) or with rainfall during store hours.[26]

d) *Number of selling days per month.*—Direct rectification of data for length of work month was, upon analysis, considered both unnecessary and unpractical. As a check upon the conclusion that its effect was negligible, number of selling days per month was tested as an independent variable.

e) *Change in volume from previous month.*—The position of a point on a static cost function is assumed to be unaffected by the position of previous observations; that is, the cost-output relation for one period is not supposed to be influenced by the output of the previous period. Our cost function, however, may not correspond precisely to this model, since rigidities of various sorts may cause the cost associated with a given output to be different when this output has been attained by an increase from the previous level than by a decrease. To examine the reversibility of the empirical cost function, the magnitude and direction of change in output from that of the previous month was tested as an independent variable.

f) *Fluctuations in business conditions.*—It might be supposed that fluctuations in general business conditions would influence costs in the present study mainly through changes in the various input prices and through variations in number of transactions and in the average value of transactions. By rectifying the data to remove the influence of changes in wage rates and prices, it was hoped to remove the greatest part of the "irrelevant" variation in costs attributable to what will be called, for lack of a better term, "the business cycle." The remaining effect of cyclical influences upon cost behavior was roughly tested by plotting the cost residuals of a graphic correlation analysis in chronological order to observe periodic fluctuations.

[26] By studying daily sales records over the five-year period, in conjunction with average noontime temperature, it was found that, on the average, sales fell off when the noontime temperature was 85° or higher in summer and 15° or lower in winter. The number of selling days with temperature above or below these critical points was therefore determined for each month of the period under study.

The effects of precipitation were studied by comparing daily sales with records of rain and snow during store hours. Snow appeared to have little or no ill effect upon sales, whereas rain was frequently associated with subnormal sales. The number of days in which it rained during store hours in each month was therefore included in the index of bad weather. Duplications between the two criteria of bad weather were then eliminated, leaving a net count of the number of selling days with disagreeable weather in each month. This involved equal weighting of days regarded as uncomfortable from each of these conditions, but no a priori or empirical evidence was available to indicate other weights.

3. *Testing the influence of independent variables.*—In the preceding section certain cost influences were tentatively selected as independent variables for the correlation analyses on the grounds that they appeared to influence cost significantly, that they could be measured and treated statistically, and that their effect upon cost could not be safely removed by rectification of the original data. Number of transactions and average value of transaction were chosen as measurement units for physical output. In addition, the following sources of cost variation were analyzed:

1. Difference between actual and anticipated volume
2. Intramonth variability in volume of transactions
3. Number of selling days having unfavorable weather
4. Number of selling days per month
5. Change in volume from previous month
6. Fluctuations in business conditions

Two criteria were applied in selecting from this list the independent variables for the least-squares multiple-correlation analysis: (1) the factor must have an independent influence upon cost not accounted for by some other variable and (2) the factor must not be highly correlated with any other independent variable.

By a preliminary graphic analysis it was possible, first, to ascertain whether any net relation existed between cost and each of the tentative causal elements; second, to define the general character of this relation; and, third, to determine the degree of intercorrelation among the independent variables.

The net relations were tested by employing each factor as an independent variable in graphic multiple-correlation analyses of total departmental cost. Introduction of the independent variables in the order of their believed importance[27] made it possible to establish by successive approximation the net relation between cost and those items which proved most important and to test the correlation between each factor and the cost variation not attributable to a more closely correlated variable.[28] A clear net relation to cost was found for both number of transactions (X_2) and average value of transaction (X_4).[29]

[27] The hierarchy was based upon opinion of executives plus closeness of simple correlation as determined by inspection of scatter diagrams. Space limitations prevent introduction of these and the multiple-correlation charts in this presentation.

[28] The order of introduction may have a significant effect upon results, hence the care with which the variables were arrayed (cf. W. Malenbaum and J. D. Black, "The Use of the Short-Cut Graphic Method of Multiple Correlation," *Quarterly Journal of Economics*, LII [1937], 66–112).

[29] "Average value of transaction" will hereafter be referred to as "average gross sale" in deference to department-store terminology.

Cost deviations from these net regression curves were plotted against each of the remaining independent variables, to determine whether the factor was significantly correlated with the residual variation in cost not accounted for by the net regressions of X_2 and X_4.

The number of selling days per month was found to have no net correlation with cost in any of the three departments after the cost variation associated with X_2 and X_4 had been removed. Although the effect of the factor upon sales volume was clear, its net effect upon cost was not; and it was, therefore, not included among the independent variables.

The percentage change in number of transactions from the corresponding month of the previous year likewise showed no net correlation with cost for any department studied and was rejected as an independent variable.

Although unfavorable weather (as measured by the number of selling days per month that were rainy or uncomfortably hot or cold) appeared to have a clear effect upon sales volume, this index showed no statistically significant net relation to cost for any department. Apparently its only effect upon cost was through its effect upon sales volume; it was therefore not used in the least-squares analysis.

Daily variability in number of sales transactions likewise failed to show a statistically significant net relation to cost in each department and was, therefore, not included among the independent variables.

Change in volume from that of the previous month showed no net relation to cost, thus roughly showing the continuity or reversibility of the cost function. No cyclical pattern was found in the cost residuals from the X_2 and X_4 net regressions. This indicated that no additional correction for cyclical changes in supply prices and wages was needed and that the observed cyclical fluctuations of cost were primarily accounted for by fluctuations in physical and dollar volume of sales.[30]

To summarize, by determining graphically the net regression of each prospective independent variable on cost, a significant relation was found between cost and two factors: number of transactions (X_2) and average gross sale (X_4). Dollar volume of sales (X_3) showed a

[30] Seasonal variation, on the other hand, did not exhaust its influence on cost by working through number of transactions and dollar value of sale. Rather it apparently exercised an independent additional influence since the cost residuals from the net regression of X_2 and X_4 had a subdued but well-defined seasonal pattern. Instead of including the seasonal factor as a third independent variable, however, the residuals were used to rectify the original cost data. This procedure is discussed under "Data Rectification."

strong gross relationship but was adequately represented in the multiple-correlation analysis by X_2 and X_4. A well-defined residual seasonal pattern of cost variation was discovered and was removed from the data by a correction index. No net relation to cost, however, was found for the following factors:

1. Change from corresponding month of previous year
2. Number of selling days per month
3. Number of selling days per month having unfavorable weather
4. Day-to-day variability in volume of transactions
5. Business conditions

4. *Determining the intercorrelation of independent variables.*—The degree of intercorrelation among the independent variables was first explored by means of scatter diagrams, which provided a sufficiently accurate indication for these purposes.

Number of transactions (X_2) and average gross sale (X_4) showed no correlation for the coat and shoe departments. For hosiery no correlation was found for the years 1932, 1933, 1934, and 1935, but for the year 1931 a clear relation appeared to exist.

The high correlation which existed between some of the rejected independent variables and number of transactions (X_2) accounts, in part, for lack of any relation between these variables and cost after the effect of volume variation had been removed.[31]

III. FINDINGS

The preceding sections have dealt with problems of collecting, rectifying, and analyzing the data in order to find the net relation between cost and output, with other influences held constant. In this section the findings of the study are presented.

Three departments of a retail store were studied: the women's medium-priced coat department, the women's hosiery department, and the women's shoe department. The cost analyzed for each department excluded general store allocated expenses and was confined to an ag-

[31] Both the number of selling days per month and disagreeable weather, for example, were clearly correlated with volume of sales, although multiple-correlation analysis revealed no additional independent effect upon cost. Likewise, the original data revealed a clear cyclical pattern in volume of transactions and in average dollar value of transaction. Yet there was no significant residual cyclical variation in corrected cost over and above that attributable to X_2 and X_4. Both number of transactions (X_2) and average gross sale (X_4) had a strong seasonal pattern. It was, in fact this high intercorrelation which precluded correction of the original data for the full seasonal influence. The residual seasonal pattern which served as a basis for the correction index was not, however, correlated with X_2 or X_4.

gregate of the monthly expenses of advertising, salespeople's salaries, other department salaries, inside delivery, outside delivery, and direct expenses. This aggregate cost, hereafter referred to as "combined cost," was studied in three forms: as a total of the monthly expense (total cost), as the average expense per unit of sale (average cost), and as the increment in total cost associated with an additional unit of sale (marginal cost).

Previous experiments in methodology have indicated that more useful and accurate estimates of cost behavior can be obtained by analyzing cost in terms of total expense for an accounting period than in terms of expense per unit of sale.[32] Cost behavior was, therefore, analyzed in terms of totals before converting the findings into average and marginal terms. Empirical cost functions were obtained by least squares multiple regression analysis of corrected monthly observations.

A. COST AS A FUNCTION OF NUMBER OF TRANSACTIONS

We shall first summarize the findings concerning cost behavior associated with variations in output as measured by number of transactions (the influence of average gross sale being allowed for) since this was accepted as the most useful measure of output.[33] The total, average, and marginal cost functions will be shown for the coat, hosiery, and shoe departments.

In order to keep absolute cost magnitudes confidential, both cost and output measures were transformed into index numbers. Number of transactions will henceforth refer to number of units of transaction index, and average and marginal costs must be understood to refer to the index unit. Average gross sale will also refer to index units rather than to dollars.

1. *Coat department*.—The following partial regression equation for total cost was obtained for the coat department:[34]

[32] Cf. Joel Dean, *Statistical Determination of Costs, with Special Reference to Marginal Cost* (Chicago: University of Chicago Press, 1936).

[33] A strong gross relation between dollar volume of sales (X_3) and both seasonally corrected and uncorrected cost was evidenced for each department. The relation of total cost appeared to be convex downward for coats and shoes (falling marginal cost), and linear for hosiery (constant marginal cost). Since dollar volume is more commonly used in store management than is physical quantity, a least-squares analysis of its relation to cost was made which confirmed these graphic results. Simple correlation was indicated because deviations from the gross regressions did not have a significant relation to average gross sale or to other factors.

[34] The following subscripts and superscripts are used in these equations: T = total cost, A = average cost, M = marginal cost; C = coat department; H = hosiery department; S = shoe department; X_2 = index of number of transactions; X_4 = index of average gross sale.

$$X^{C}_{T} = 16.835 + 1.052X_2 - .00194X_2^2 .$$

It shows total cost increasing in a convex curve which rises at a declining rate as physical volume increases, as portrayed in Chart I.

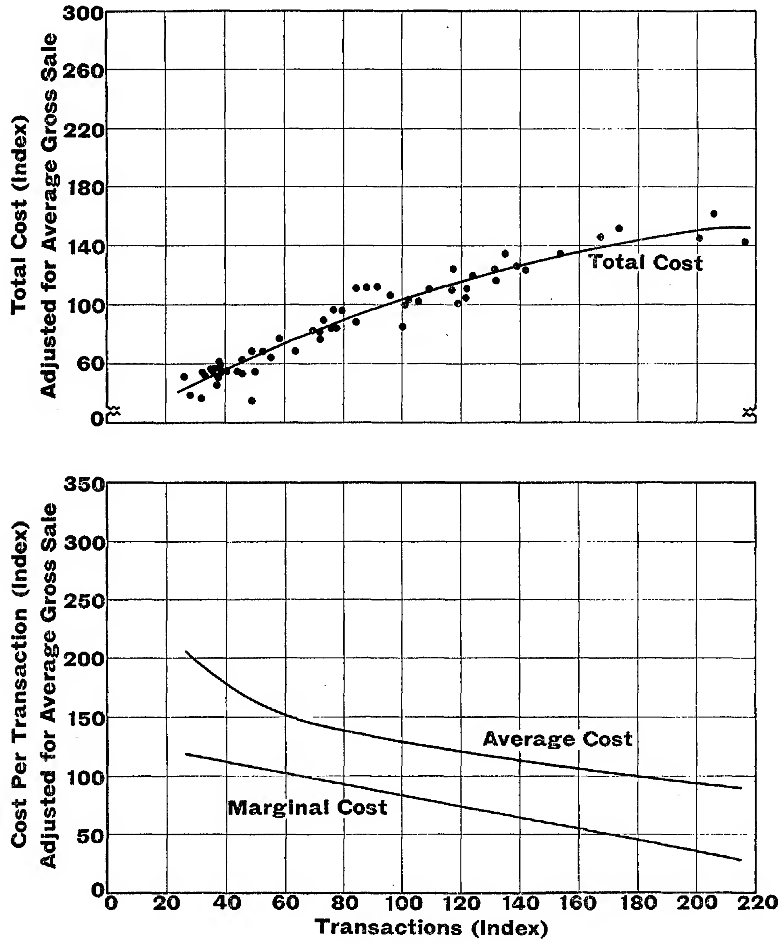


CHART I
COAT DEPARTMENT: TOTAL, AVERAGE AND MARGINAL COMBINED COST DERIVED FROM PARTIAL REGRESSION ON TRANSACTIONS

The behavior of average cost was determined by conversion of the corresponding total cost function rather than by direct correlation analysis of unit cost observations. It shows transaction cost declining in a hyperbolic curve as physical volume increases, as may be observed in Chart I, and seen from the following equation:

$$X^{C}_{A} = 1.052 + 16.835/X_2 - .00194X_2 \, .$$

The marginal cost per transaction unit declines at a constant rate, as physical volume increases, in a function described by the following equation:

$$X^{C}_{M} = 1.052 - .00388X_2 \, .$$

2. *Hosiery department.*—For the hosiery department the least-squares equation for the net relation between cost and number of

TABLE 1

HOSIERY DEPARTMENT: ESTIMATED TOTAL, AVERAGE, AND MARGINAL COST* FOR VARIOUS UNITS OF TRANSACTIONS (X_2) WITH THE INDEX OF AVERAGE GROSS SALE CONSTANT AT ITS MEAN

Transactions (Index)	Estimated Total Cost (Index)	Estimated Average Cost (Index)	Estimated Marginal Cost (Index)
60	76.374	1.273	.347
80	83.314	1.041	.347
100	90.254	.903	.347
120	97.194	.810	.347
140	104.134	.744	.347
160	111.074	.694	.347
180	118.014	.656	.347
200	124.954	.625	.347
240	138.834	.578	.347
280	152.714	.545	.347
320	166.594	.521	.347

* Derived from the equations:

$$X^{H}_{T} = 55.554 + .347\,X_2$$
$$X^{H}_{A} = X^{H}_{T}/X_2 = .347 + 55.554/X_2$$
$$X^{H}_{M} = .347$$

transactions, while holding the average gross sale at its mean, is

$$X^{H}_{T} = 55.554 + .347X_2 \, .$$

The graph of the equation, which indicates that total cost rises at a constant rate as number of transactions increases, is shown in the upper half of Chart II. Inspection of the latter figure reveals that the observations upon which the curves are based are so unevenly distributed that the relation was well established only between the transactions index figures of 80 and 170, and only tentatively defined for the range between 170 and 300.

The equation for average cost per sales transaction as related to number of transactions, when the average gross sale is held constant, is

$$X^{H}_{A} = .347 + 55.554/X_2 \, .$$

In tabular form this function is presented in Table 1. Examination of

the table shows that average cost per transaction unit falls at a declining rate as physical volume of sales increases and tends to approach a constant at the extreme range of cost observation.

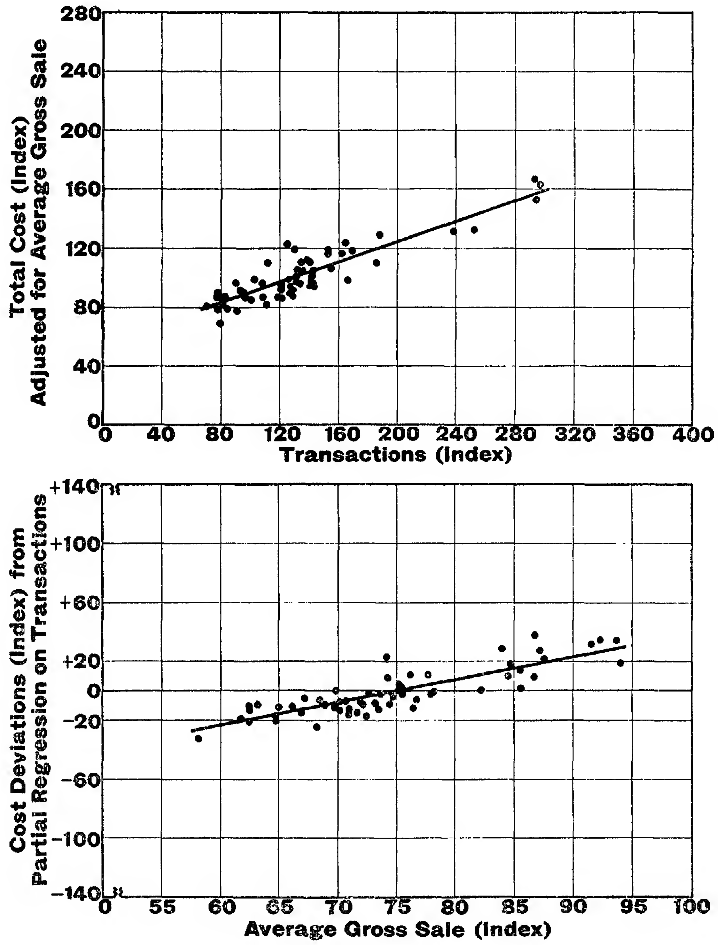


CHART II

HOSIERY DEPARTMENT: PARTIAL REGRESSIONS OF TOTAL COMBINED COST ON TRANSACTIONS AND AVERAGE GROSS SALE

The marginal cost of one additional hosiery transaction was found to be constant at .347.

3. *Shoe department.*—For the shoe department the net relation between total cost and number of transactions (X_2), when the average value of transaction is held constant at its mean, is depicted by the following partial regression equation:

$$X^{S}{}_{T} = 32.137 + .925X_2 \,.$$

The graph of this function (Chart III) indicates that total selling cost of the department increases at a constant, although not proportionate, rate as the physical volume of sales rises.

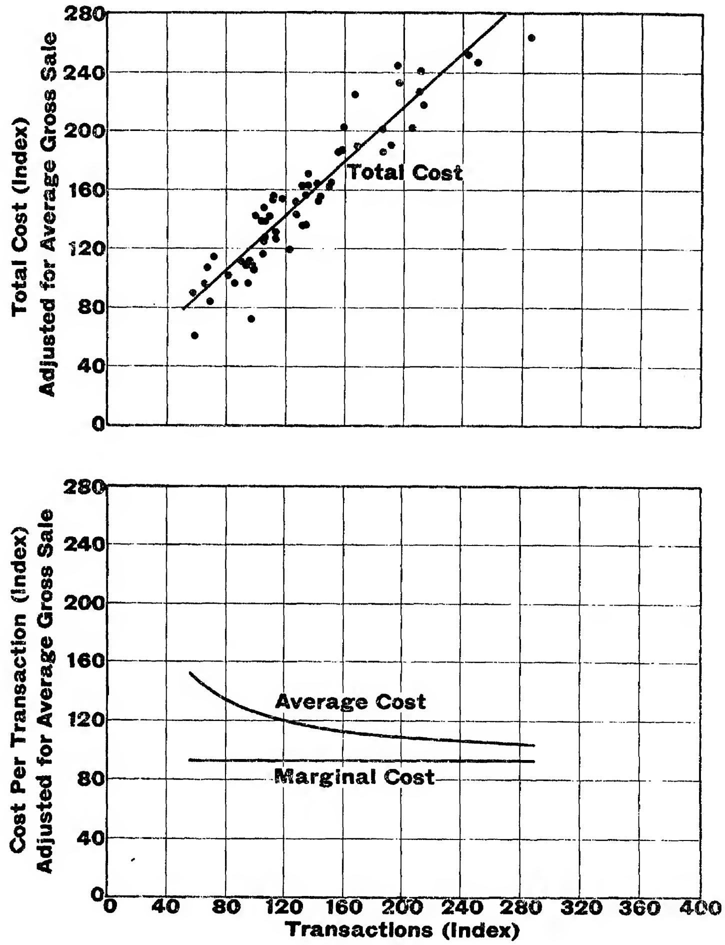


CHART III

SHOE DEPARTMENT: TOTAL, AVERAGE AND MARGINAL COMBINED COST DERIVED FROM PARTIAL REGRESSION ON TRANSACTIONS

The average cost per transaction unit for women's shoes declines, as would be expected, when the number of transactions increases. This relation, when average value of sale is constant at its mean, is shown in the following equation:

$$S^S{}_A = .925 + 32.137/X_2 .$$

The graph of the function is shown in Chart III, where it may be contrasted with the total cost function.

The estimate of the marginal cost of an additional transaction index unit for the women's shoe department appears to be constant, throughout the observed range of physical volume, at .925.

B. COST AS A FUNCTION OF AVERAGE GROSS SALE

The findings for cost functions when number of transactions is accepted as the index of the output (with average gross sale held constant) which were presented in the preceding section need to be supplemented by examination of the influence of another aspect of output. The significance of the net relation between the average gross

TABLE 2

BUDGET COAT DEPARTMENT: ESTIMATED TOTAL, AVERAGE, AND MARGINAL COST* FOR VARIOUS MAGNITUDES OF THE INDEX OF AVERAGE GROSS SALE (X_4) WITH INDEX OF TRANSACTIONS CONSTANT AT ITS MEAN

Average Gross Sale (Index)	Estimated Total Cost (Index)	Estimated Average Cost (Index)	Estimated Marginal Cost (Index)
50	75.292	1.506	.787
60	83.162	1.386	.787
70	91.032	1.300	.787
80	98.902	1.236	.787
90	106.772	1.186	.787
100	114.642	1.146	.787

* Derived from the equations:

$$X^C{}_T = 35.942 + .787\, X_4$$
$$X^C{}_A = X^C{}_T/X_4 = .787 + 35.942/X_4$$
$$X^C{}_M = .787$$

sale and cost indicates that output should be considered as two dimensional and may increase either in the direction of more transactions or greater average size of the transaction. Accepting size of transaction (average gross sale) as the measure of output, and holding number of transactions constant at its mean, we obtain the results summarized in this section.

1. *Coat department.*—The net functional relation between total cost of the coat department and average gross sale (with number of transactions constant) is shown by the following least-squares partial regression equation:

$$X^C{}_T = 35.942 + .787X_4 .$$

Total cost tends to increase at a constant rate with the size of the av-

erage gross sale. Table 2, which was obtained by substituting in the above regression equation, shows this relation in the form of estimated total cost for various values of average gross sale.

The average cost per index unit of average gross sale, as average gross sale varies and number of transactions remains constant, is shown by the following expression:

$$X^{C}_{A} = .787 + 35.942/X_4 .$$

The cost per unit increase in the monthly average value of transaction is constant at .787 over the observed range.

2. *Hosiery department.*—For the hosiery department the equation obtained for the net relation between total costs and average gross sale, when the effect of number of transactions is allowed for, is

$$X^{H}_{T} = -13.889 + 1.557X_4 .$$

TABLE 3

HOSIERY DEPARTMENT: ESTIMATED TOTAL, AVERAGE, AND MARGINAL COST* FOR VARIOUS MAGNITUDES OF THE INDEX OF AVERAGE GROSS SALE (X_4) WITH INDEX OF TRANSACTIONS CONSTANT AT ITS MEAN

Average Gross Sale (Index)	Estimated Total Cost (Index)	Estimated Average Cost (Index)	Estimated Marginal Cost (Index)
50	63.961	1.279	1.557
60	79.531	1.326	1.557
70	95.101	1.359	1.557
80	110.671	1.383	1.557
90	126.241	1.403	1.557
100	141.811	1.418	1.557

* Derived from the equations:

$$X^{H}_{T} = -13.889 + 1.557\,X_4$$
$$X^{H}_{A} = X^{H}_{T}/X_4 = 1.557 - 13.889/X_4$$
$$X^{H}_{M} = 1.557$$

The curve of this equation found in the lower half of Chart III shows that cost increases with the size of the average gross sale at a constant rate. As in the case of number of transactions, the wide dispersion of the observations casts some doubt on the validity of a linear relation.

The average cost per index unit of average gross sale, when the number of transactions is held constant, is described by the following equation:

$$X^{H}_{A} = 1.557 - 13.889/X_4 .$$

The tabular representation is shown in Table 3.

The additional cost of increasing average gross sale by one index

unit for the month was found to be 1.557.

3. *Shoe department.*—The functional relation of total cost and average gross sale, with transactions constant, for the shoe department, is shown by the equation:

$$X^S{}_T = -23.427 + .837X_4 .$$

Total cost increases at a constant rate with the value of average gross sale, when physical volume is held constant at its mean.[35]

The average cost per unit of average gross sale, with physical volume constant at its mean and size of gross sale varying, is shown by the following expression:

$$X^S{}_A = .837 - 23.427/X_4 .$$

The cost of a dollar increase in average value of transaction is constant at .837.

C. COST AS A FUNCTION OF NUMBER OF TRANSACTIONS AND AVERAGE GROSS SALE

The findings concerning the combined effect upon cost of the two aspects of output—number of transactions and average gross sale—will be briefly presented in this section.

For the coat department the combined effect of number of transactions (X_2) and average gross sale (X_4) on total cost is shown by the following multiple-regression equation:

$$X^C{}_T = -35.440 + 1.052X_2 - .00194X_2{}^2 + .787X_4 .$$

The standard error of estimate, adjusted for degrees of freedom, was 7.901. The close relation of cost to the independent variables is shown by the value of the coefficient of multiple correlation, .980, and by the coefficient of multiple determination which indicates that about 96 per cent of the variance in rectified cost was accounted for by the independent factors.

The combined effect of number of transactions (X_2) and average gross sale (X_4) upon total cost of the hosiery department is described by the following equation:

$$X^H{}_T = -60.764 + .347X_2 + 1.557X_4 .$$

The confidence which can be placed in the relation described above

[35] Graphic correlation analysis paralleled the least-squares findings by revealing the same type of net relation, although the uneven and scattered distribution of the observations from which the relations were obtained somewhat restricts the reliance to be placed in these functions.

is indicated by the standard error of estimate of 8.140, the coefficient of multiple correlation of .957, and the coefficient of multiple determination of .91.

For the shoe department the combined effects of number of transactions (X_2) and average gross sale (X_4) upon total cost is shown by the following multiple-regression equation:

$$X^S{}_T = -146.776 + .925X_2 + .837X_4 .$$

Provided that the basic conditions of the sampling period remain unchanged, considerable confidence can be placed on the above relations, as is evidenced by a standard error of estimate of 16.624, a coefficient of multiple correlation of .965, and a coefficient of multiple determination of .93.

IV. INTERPRETATION OF FINDINGS

The findings of an empirical study of the individual firm should constitute some kind of an evaluation of what theorists have been saying with regard to its economic behavior. However, such an appraisal would imply that the situation described by the statistician is strictly comparable to that generally postulated by the theorist. The actual situation investigated, in fact, deviates from the theoretical norm in two important respects, namely, (*a*) by the inclusion of selling costs in cost of production and (*b*) by the failure to purge the statistical data of all sources of dynamic nonconformity to the static model. Some brief comments will be made on these limitations of the findings in Sections A and B before an attempt is made in Sections C and D to explain and rationalize the results observed.

A. INCLUSION OF SELLING COSTS

As was mentioned above, the costs examined in this study were not confined to what is usually designated as cost of production proper. A priori, of course, there are equally good, if not better, reasons for supposing that marginal selling expenditure increases with intensified selling activity as there are for expecting that marginal cost of production rises with output. First, with more intensive utilization of fixed equipment returns per unit of input presumably decrease. Second, each additional unit of promotional cost has diminishing effectiveness since it becomes necessary, in order to increase sales, to detach customers with increasing degrees of personal affiliation and loyalty

to rival firms, or to make present customers spend an ever increasing share of their incomes.[36]

Although our findings may be strongly influenced by the presence of costs designed to modify the existing demand functions, the relation of such costs to sales may not be revealed in our data. A spurious correlation between selling expenditure and sales could result from independent fluctuations in demand and from market-sharing, since it obviously is not possible to attribute all shifts in the store's demand functions during the period of analysis to its promotional activities. It is conceivable that selling costs in a mature competitive retail market have the effect of maintaining the store's share of the total demand —a demand which shifts seasonally and cyclically in response to changes in custom, in tastes, and in income.

It would be impossible to establish the relation between number of sales transactions and expenditure on advertising without eliminating variations in sales resulting from varying effectiveness of advertising, from changes in demand, and from other irrelevant factors. The same holds true of other elements of promotional expenditure. Consequently, our results cannot be interpreted as measuring the influence of the diminishing effectiveness of promotional expenditures as a firm's market is expanded at the expense of rivals.

The inclusion of both selling costs and pure costs of production in the empirical cost function, furthermore, indicates that the product, a unit of sales service, is not entirely homogeneous, since selling activity which increases sales may be considered to have enhanced the utility of commodities offered. However, there may be considerable stability in the proportion of selling costs to total costs.[37]

An additional reason for suspecting that the unit of output, sales service, may not be homogeneous over the observation period and may

[36] "When a field is fertilized or a pig fed, the profit per dose decreases with the quantity of doses, and it is time to stop when profit and dose have the same value. Similar phenomena of fatigue appear when we feed a market, which may also be considered as a kind of living being . . . . " (Ivar Jantzen, "Increasing Returns in Industrial Production," *Nordisk tidsskrift for teknisk økonomi*, March-June, 1939, p. 6).

[37] It was not possible to isolate the promotional or selling constituent in the various accounting categories of cost. Nevertheless, in the coat and hosiery departments the behavior of newspaper advertising costs (which approximate pure promotional costs) indicated the existence of a proportional relation to output. In a graphic correlation analysis the functions appeared to be linear and to pass through the origin, although the scatter was wide, particularly for high outputs. An ill-defined curvilinear function was found for the shoe department. It seems reasonable, moreover, to expect substantial constancy of store standards over time, which would further buttress the assumption of approximate constancy of promotional components of cost. The results of this partial analysis are not conclusive, however, because of the failure to stabilize demand and other impinging conditions.

be correlated with the physical transaction rate is the fact that sales service per transaction is likely to diminish during busy seasons. As a consequence of this, the decline of marginal cost as the volume of transactions increases may merely reflect deterioration of sales service standards.[38] If this is the case, and if this deterioration has a tendency to decrease future marginal revenue, the management may regard falling marginal cost as a danger signal, indicating that sharply rising marginal promotional expenditure may be necessary in the future.[39]

B. STATISTICAL SOURCES OF NONCONFORMITY

Besides the problems involved in promotional activity, there also exist certain statistical sources of possible nonconformity of the empirical situation to the theoretical model in addition to those mentioned in the sections dealing with methodology.

It is possible that, had a smaller observation unit been chosen (e.g., days or weeks instead of months), a more conventional cost function would have been obtained. If uncorrected data had been used, the expectation would be that the longer the time unit chosen the less would be the slope of the marginal cost curve, since marginal costs are a function of the time period allowed for adjustment to changed rates of output. If rectification of the data has been successful, however, adjustments of a long-period or quasi-long-period character have been eliminated, so that there should be no discrepancy between a function based on monthly findings and one based on a shorter time period.

From the purely technical viewpoint, it should probably be mentioned that the total cost functions are not defined beyond question as being straight lines or parabolas. The unexplained scatter of observations is great enough to permit a cubic of the traditional form to be fitted in each case. However, the curvature would be so slight as to be insignificant from a managerial viewpoint, so that it could

[38] The corrected data employed are intended, however, to represent a fairly uniform quantity of sales service. Day-to-day and hour-to-hour variation in sales service was probably averaged out by taking monthly totals rather than daily or weekly totals, and the effect of seasonal variation in sales service was at least partially removed by means of the index of seasonal variation.

It is possible that the form of the cost function obtained was altered by the seasonal correction of the cost data. This appears unlikely, however, because the influence of volume on cost was removed before constructing the seasonal index, by the process described in Sec. II, 4.

[39] Since the cost components entering into sales service are not independent of marginal revenue, it is therefore possible that declining marginal cost indicates that the department has passed the output which will maximize revenue over a period of time extending more or less indefinitely into the future.

scarcely affect any economic conclusions which might be derived from the linear and parabolic functions. In each instance a higher-order function than that selected was fitted and subjected to critical ratio tests, which indicated that the more complex function did not fit the data significantly better than that chosen.

C. EXCESS CAPACITY

If the plant of these three departments were systematically underutilized during the observation period, marginal costs might appear to be constant or falling over the whole length of the curve because of an inadequate range of observations.

Overbuilding as a result of errors of optimism in estimating the position and inclination of demand functions is not a plausible explanation for overcapacity, since department-store layouts are relatively easy to modify when such errors become apparent. This explanation, moreover, does not take account of the possibility that selling expenditures may eventually shift the demand functions so that capacity is fully utilized.

The equilibrium position of a department may be one of apparent excess capacity only. The fact that the magnitude of selling expenditures affects both the cost function and the demand function results in complex interrelations that may yield equilibrium at a point on the demand surface that appears to represent underutilization of capacity. Delays in the responsiveness of demand to various kinds of promotional expenditures, together with complexities of polyperiodic production, may accentuate this condition. Store prestige may, furthermore, force some departments to operate at less than their departmental optimum capacity. An overelaborate layout and stock in particular departments could also result from the effort of the store to maintain comparable standards of sales service, variety of merchandise, display, etc., in all departments of the store. Store "good will" may thus necessitate a minimum size for the departments studied, greater than that justified by the average position and slope of their demand functions during the period examined.

We have no evidence that the observed outputs were those which would have maximized profits under these complicated conditions. In view of the complexity of the relations between cost and revenue, it is not surprising that this store conformed to the general pricing policy of department stores which is based on the "cost plus a given percentage" principle rather than on an attempt to calculate optimum price on the basis of interrelated cost and revenue functions. The

"cost-plus" policy would yield the price and output which was "optimum" for any momentary situation only accidentally and might result in a systematic underutilization of plant which would confine observations to the constant section of the marginal curve.

D. TECHNICAL EXPLANATIONS FOR CONSTANT MARGINAL COSTS

A variety of technically plausible explanations can be given for findings of constant marginal costs in the hosiery and coat departments and falling marginal costs in the shoe department. Marginal costs may be constant if it is possible to "sectionalize" the activities of a firm; that is, if the factors employed can be grouped into small operating entities, each of which can be utilized to equal advantage, and each of which represents the optimum combination of factors.[40] In practice, of course, a plant cannot be completely sectionalized, since some factors, if only the management and organization, are fixed in the short run. As long as fixed factors are present, the law of diminishing returns will cause costs to rise over some portion of the marginal cost curve.

Second, the declining phase of the marginal cost curve will be particularly important if the fixed factors are not completely adaptable to varying inputs of the variable factors and if, in addition to "lumpiness" of the fixed factors, there are organizational economies available when larger amounts of the variable factors are used. The inclination of the curve will depend on the degree of technical "flexibility" to be found. An inflexible plant would have a U-shaped marginal cost function with a sharply defined minimum; greater flexibility would be present if outputs less or greater than the optimum could be obtained without entailing rapidly rising marginal cost.[41]

It is probable that the type of marginal cost function found in empirical studies can be explained by investigating (1) the degree to which sectional divisibility of the plant has been attained and (2) the degree of technical rigidity which exists.

In a department store technical rigidities are probably not so great as would be found in a manufacturing enterprise. The fixed elements of cost include, from the standpoint of the individual depart-

[40] Cf. M. F. W. Joseph, "A Discontinuous Cost Curve and the Tendency to Increasing Returns," *Economic Journal*, XLIII (September, 1933), 390–98.

[41] See a recent article by Mr. George Stigler, entitled "Production and Distribution in the Short-Run" (*Journal of Political Economy*, XLVII [1939], 305–28), for a discussion of the concepts of adaptability and flexibility. Mr. Stigler has pointed out that technical flexibility may be deliberately provided by building a plant in a way which will maximize profits over time through efficient operation over the expected range of outputs, even if this involves the sacrifice of the optimum technical adjustment for any one output.

ment, the general standards and reputation of the store, the location, in the present instance the size of the departmental layout, and to a certain extent the services of the buying and managerial staff. However, it is probable that buying and managerial service varies with output, even when no recognition of such variation is given in salaries or number of people employed. To this relatively small component of fixed factors might be added units composed of the optimum combination of selling service, advertising, delivery expense, etc. The degree of "sectionalization" thus attained might offset, for considerable variations in output, the tendency to diminishing returns. Moreover, since there are few highly specialized input factors in a technical sense the degree of flexibility is probably great.

In conclusion, since the demand and cost functions are not independent when promotional expenditures are being made, it would be necessary in order to get a determinate solution to the problem of the optimum price and output not only to separate promotional costs from cost of production proper or to determine a functional relation between them but also to determine (1) the relation between selling costs and the demand function, (2) the time lag between the incurring of the selling expenditure and its effect on demand, (3) the relation between the output-price adjustment in one period and the demand function in successive periods, (4) the interrelation among the demands for various products both within and outside this department (e.g., "loss leader" and "ensemble" buying), and (5) the reaction of rivals to the firm's price, output, and promotion policies.

AN ACCELERATION PRINCIPLE FOR COMMODITY STOCKHOLDING AND A SHORT CYCLE RESULTING FROM IT

J. TINBERGEN
Rotterdam School of Economics

I. COMMODITY STOCKS IN THEORY AND STATISTICAL RESEARCH

FLUCTUATIONS in stocks of commodities are neglected in static theory, but they play an interesting role in economic dynamics. Stocks are, at any given moment, one of the "inheritances" of the past, and thus their influence on other phenomena offers an interesting example of a dynamic relation which various business-cycle theorists have introduced into their picture of causations. Stocks represent overproduction if they are abundant; while, if they are small, they represent scarcity and lead to "bottlenecks" and the sort. Investment may take the form of deliberate accumulation of stocks or of stocks of goods in process considered as "working capital," and opinions have been given on the influence of interest rates on stocks. These few examples seem to be sufficient to recall to the reader the interest given to commodity stocks in dynamic theory.

As far as statistical observation is concerned, the importance of raw material stocks in the hands of dealers for the formation of prices is a recognized fact. The fluctuations in prices of agricultural raw materials can be explained to a large extent by the fluctuations in "quantity available" which is the sum total of carry-over plus new crop. Visible stocks may be taken instead of quantities available, since they show about the same fluctuations.

For the explanation of the fluctuations in mineral raw material prices the fluctuations in stocks are also relevant, as is shown by recent investigations for tin, copper, and iron. Figures on other types of stocks—stocks of semifinished and finished articles or of raw materials in the hands of their consumers—are not available in great abundance, but they are steadily increasing. The factors determining their fluctuations have as yet hardly been studied systematically by statistical methods. The present paper is a modest contribution in this field.

II. FACTORS DETERMINING THE FLUCTUATIONS IN STOCKS

It seems useful to make a distinction between stocks of goods which are produced or bought deliberately and stocks of goods the holding or production of which is beyond human control. The latter category consists chiefly of vegetable raw materials in the first hand. This may be the farmer, but in many cases dealers in producing countries take their place; more exactly we mean by the first hand those that bear the risks of production fluctuations. To a large extent the stocks of these goods in the second hand are of the first-mentioned category and so are all stocks of other commodities. In the following considerations only this first category of deliberately regulated stocks will be considered.

There are various technical reasons for holding such stocks:

1. The necessity of exposing to buyers the assortment available
2. The necessity of linking up continuous production with discontinuous consumption or discontinuous production with continuous consumption
3. The necessity to meet irregularities either in production or in deliveries

Apart from these technical reasons there are reasons of an economic character:

4. Stocks may be held as a form of investment, especially if interest rates are low or money is depreciating or expected to do so ("the flight into goods")
5. If price gains are expected for the special commodity considered

Since various attempts to explain the fluctuations in stocks have brought us to the conviction that the technical reasons for holding stocks are of preponderant importance, we shall go into some more detail concerning points 1, 2, and 3 above. What will the fluctuations be that are the consequences of these principles of stockholding?

Stocks for exposition need only show changes if the number of models or of qualities of a certain product changes. In general, this type of change will be slow and continuous; these changes are "structural."

Stocks for meeting regular or irregular interruptions of either production or consumption all have in common that they should be proportional to production or consumption. This proportionality leads to an "acceleration principle" in that demand for stocking will be proportional to the rate of increase in production or consumption, just as in the ordinary acceleration principle the demand for durable investment goods is assumed to be proportional to the rate of increase in production of consumers' goods.

In order to draw some conclusions from this statement, we have

to go into some more detail. Considering the various stages between raw material imports and finished goods consumption, we may distinguish for each of them an "intake" and an "output." The intake of the first stage is identical with raw material imports; the intake of each further stage is identical to the output of the previous stage, and the output of the last stage equals final consumption. For each of the stages separately the stock may be held proportional to intake or to output. Presumably the latter is the more common behavior. Looking at all stages together, this would, however, mean that total stocks are proportional neither to intake for the first stage nor to output for the last stage, but approximately proportional to something in the middle.

Since statistical evidence is seldom available for all stages separately, we want to simplify the picture and only to introduce intake for the first stage (imports) and output for the last stage (consumption). Alternatively, we shall assume stocks to be proportional to (1) consumption (Case I) and (2) imports (Case II).

Case I.—Denoting stocks at end of unit time period t by s_t, imports during that period by u_t, and consumption during that period by u'_t, we have

$$u_t = s_t - s_{t-1} + u_t' . \tag{1}$$

Assuming

$$s_t = \alpha\, u_t' , \tag{2}$$

we get

$$u_t = \alpha(u_t' - u'_{t-1}) + u_t' . \tag{3}$$

Case II.—In this case we assume that

$$s_t = \alpha u_t , \tag{2'}$$

by which we get

$$u_t = \alpha(u_t - u_{t-1}) + u_t' . \tag{3'}$$

We did not yet discuss whether the tendency to hold stocks proportional to u_t' or u_t works without any lag or with some lag. The existence of a lag seems more probable. Strictly speaking such a lag has already been assumed to exist, since s_t relates to the end of the time unit and u_t to the period itself. The longer the period, the longer is the lag involved. Of course, it may also be that relation (2′) takes the form

$$s_t = \alpha u_{t-\theta} , \tag{2''}$$

which can only approximately be brought into the form (2′) by a change of time units into new units $2\theta + 1$ as large as the old ones.

Relations similar to (3) but in many respects more complicated have been discussed at length by B. A. Chait.[1] Chait does not, in his book, consider the possibility of a lagged relation[2] but considers separately all the stages.

III. FIGURES ON THE ACCELERATION PRINCIPLE FOR STOCKS

Before continuing our argument, we may first show some statistical evidence concerning the "acceleration principle for stockholding."

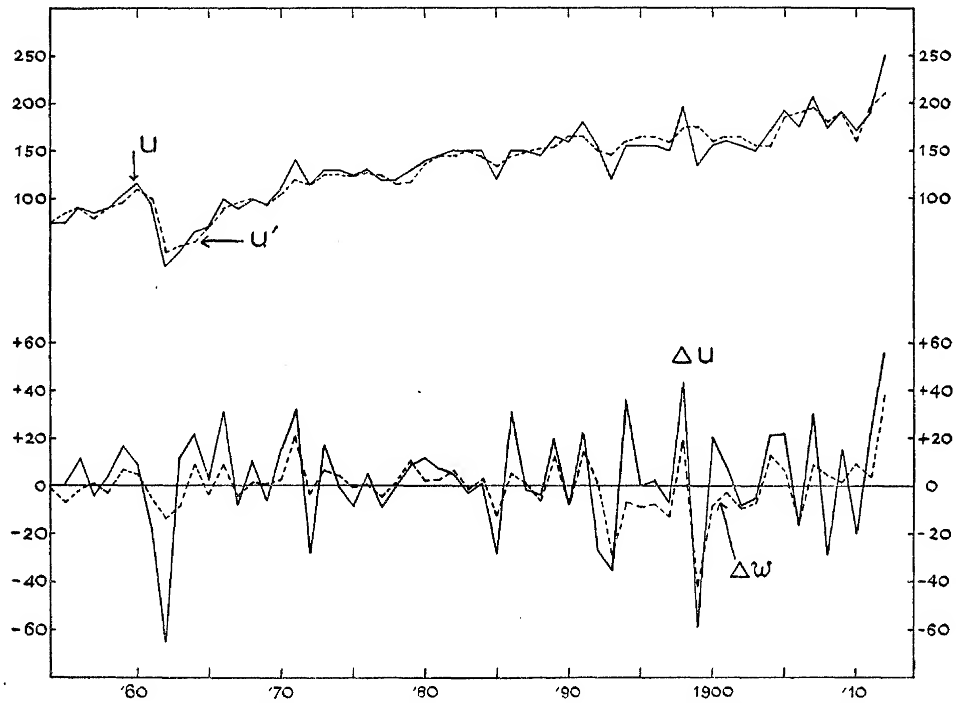


FIG 1.—United Kingdom: Cotton. u: imports, u': consumption, 10^7 lbs. w = rate of increase in stocks $= u - u'$.

A long series of figures is available for cotton imports into the United Kingdom and for cotton consumption by manufacturers. The difference evidently is the increase in stocks. Comparing this difference with the rate of increase in imports as compared in the previous year, we find a striking similarity (Fig. 1, lower graph). There is also a rather good correlation with the rate of increase in consumption,

[1] *Les Fluctuations économiques et l'interdépendance des marchés* (Brussels, 1928), esp. pp. 63–77.

[2] He does so, however, in a publication shortly to appear: *Le Dynamisme des marchés: la loi de divergence* (Paris: Hermann & Cie).

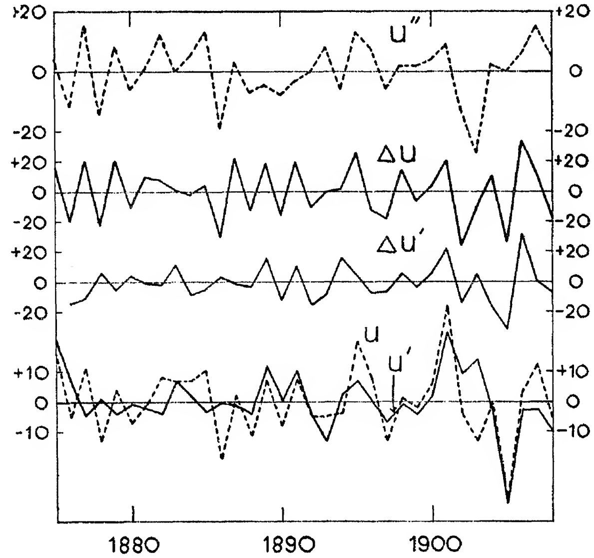


FIG. 2.—United Kingdom: Sugar. U: imports, u': consumption, 10^5 cwts. $u'' = u - u'$. All series: deviations from nine-year moving averages.

but this one seems to be less good than the previous one. This means that our formula (2′) applies if a year is chosen as the time unit. The

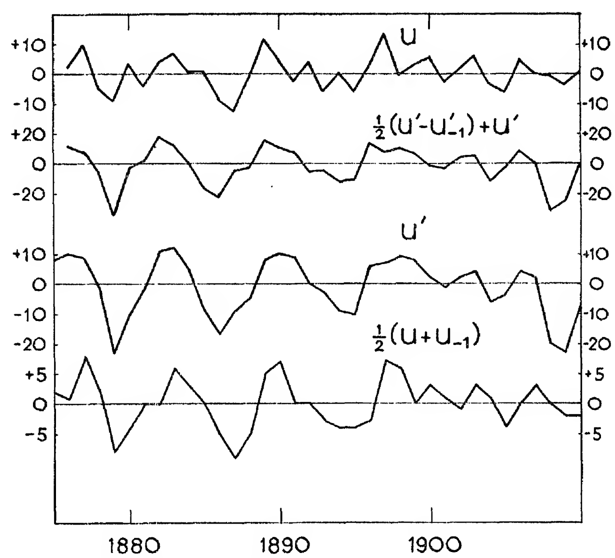


FIG. 3.—United Kingdom: Wood imports and building activity, 10^5 loads. u: imports of wood. u': employment of carpenters and joiners. All series: deviations from nine-year moving averages.

value of α seems to be approximately $\frac{1}{2}$. Since there is a danger (as we shall discuss in Sec. VI) of spurious correlation, the value of this coefficient will be reconsidered later.

The same sort of data are available for sugar and are represented in Figure 2. For comparison also the rate of increase in consumption has been added. Although the correlations are decidedly less good than in the case of cotton, they do not seem to be entirely meaningless.

A somewhat more indirect test can be made for wood and timber, where neither consumption nor stock data are available, but where it may be assumed that consumption shows fluctuations parallel to those

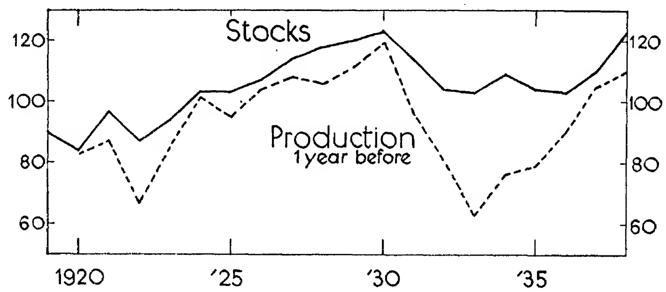


FIG. 4.—United States. Stocks: Index of stocks of finished goods, Department of Commerce. Production: Index of industrial production, F.R.B. 1923–25 = 100.

in employment of carpenters and joiners. With the help of Figure 3 two comparisons may be made, corresponding to formulas (3) and (3′). For convenience α has been taken equal to $\frac{1}{2}$. Formula (3) then reads

$$u_t = \tfrac{1}{2}(u_t' - u'_{t-1}) + u_t' \, ,$$

and formula (3′) may be written

$$u_t' = \tfrac{1}{2}(u_t + u_{t-1}) \; .$$

It would seem that (3) gives a closer fit for the second half of the period and (3′) for the first half.

Finally, Figure 4 provides some data on stocks of finished goods and retail stocks in the United States. In these cases, also, there is a clear tendency to a lagged parallelism with such indices of activity as the volume of production and the volume of retail sales.

IV. A CONSEQUENCE OF THE ACCELERATION PRINCIPLE FOR STOCKING: THE TENDENCY TO A SHORT CYCLE

We may now consider in some more detail the consequences of our acceleration principle for the movements in imports. Since our alterna-

tive formulas do not lead to quite the same conclusions, we shall consider both in turn.

Case I.—Formula (3) at once gives the fluctuations of imports. It does not, however, provide us with a simple means to see, in general, how these fluctuations differ from those in consumption. We may calculate a standard case by assuming that u' is first constant $= u_0'$, then suddenly jumps to a new value u_1'. In a sense, all changes in u' may always be considered as repetitions of this standard case. The values in u_t will be:

	TIME							
	0	1	2	3	4	5	6	7
u' . . .	u_0'	u_0'	u_0'	u_0'	u_1'	u_1'	u_1'	u_1'
u . . .		u_0'	u_0'	u_0'	$u_1' + \alpha(u_1' - u_0')$	u_1'	u_1'	u_1'

The result is represented graphically in Figure 5, where also the deviations between u and u' have been indicated. These may be characterized —if one likes—as one single short wave.

There will be more than one single wave if we consider a system

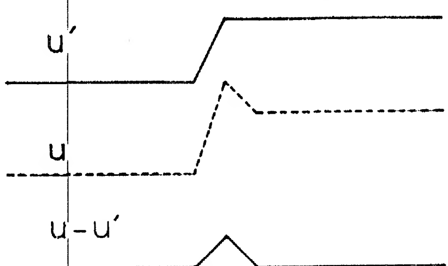


FIG. 5.—Course of consumption u' and imports u, if u' shows one jump and formula (3) is valid.

of more than two markets, for example, if the stage of importation (u) would be preceded by one, say, of production in another country (v), for which we have an equation

$$v_t = \beta(u_t - u_{t-1}) + u_t\,,$$

quite analogous to (3). In the simple case now considered this yields the following values for v_t :

0	1	2	3	4	5	6	7
		u_0'	u_0'	$u_1'+\alpha(u_1'-u_0')+\beta(1+\alpha)(u_1'-u_0')$	$u_1'-\alpha\beta(u_1'-u_0')$	u_1'	u_1'

Case II.—The nature of the movements resulting from formula (3′) is different. This may be seen best by assuming u_t' to be con-

stant. If we measure u_t in deviations from this constant value u_t', we get, instead of (3′),

$$u_t = \alpha(u_t - u_{t-1}) \; , \tag{4'}$$

which may be written

$$u_t = - \frac{\alpha}{1-\alpha} u_{t-1} \; . \tag{5'}$$

From this formula it follows that u_t will show movements even if u_t' will not. Instead of one single wave, as found for Case I, we will have an endless series of short waves of a period of two time units. These cycles will be damped if $\alpha/(1-\alpha) < 1$ or $\alpha < \frac{1}{2}$, whereas they might even be antidamped for $\alpha > \frac{1}{2}$.

Similar results are obtained if we choose somewhat different approaches instead of formula (3′). Such an approach—somewhat more satisfactory from the theoretical viewpoint—might be the replacement of $u_t - u_{t-1}$ by the rate of increase calculated with the help of shorter time periods. Suppose we take our time unit three times as small as before. If then we assume that the rate of increase in intake $u_{t-1} - u_{t-2}$ determines the addition to stocks in unit t, we get, in analogy to (4′),

$$u_t = \alpha'(u_{t-1} - u_{t-2}) \; .$$

The lag is now $1\frac{1}{2}$ new time units which is equal to $\frac{1}{2}$ old unit and therefore corresponds exactly to our former case.

Since now the u's relate to time periods three times as small as before, they are in general three times as small as before. The rate of increase per old time unit will be, in the average, nine times as large as the rate of increase per new time unit, since not only the figures u but also their distances are now three times as small.

In order, therefore, to have an exact "translation" of our former case into the language of the better approximation, we have to choose $\alpha' = 3\alpha$. Thus we get

$$u_t = 3\alpha(u_{t-1} - u_{t-2}) \; . \tag{6}$$

The solution of this difference equation may be found by putting $u_t = Cx^t$, where C and x are constants, C is arbitrary, and x has to be determined from

$$x^2 - 3\alpha x + 3\alpha = 0 \; , \tag{7}$$

being the characteristic equation to (6). The values of x satisfying (7) are

$$x_{1.2} = \frac{3}{2}\alpha \pm \sqrt{\frac{9}{4}\alpha^2 - 3\alpha} \ .$$

The corresponding movements of u will be cyclic if

$$\frac{9}{4}\alpha^2 - 3\alpha < 0$$

or

$$\frac{3}{4}\alpha < 1$$

$$\alpha < \frac{4}{3} \ .$$

There will be damping if $3\alpha < 1$ or $\alpha < 1/3$.

The limits are somewhat different, but qualitatively the same phenomena present themselves. The period is now no longer a constant but varies with α. For some values of α they are indicated in Table 1 (in "old" time units). In the neighborhood of $\alpha = 1/3$ about the same periods are found as before.

TABLE 1

α	Period	α	Period	α	Period
1/6 . . .	1.7	1/2 . . .	2.3	1 . . .	4
1/3 . . .	2	2/3 . . .	2 2/3	4/3 . . .	∞

Other alternatives might be tried. Instead of a simple lag a distributed lag could be introduced. The difference might be calculated over still another interval, etc. We do not think this is worth while as long as no more exact statistical data have been investigated. For this sketch we stop at this second approximation.

Case I as well as the alternatives for Case II considered above have one feature in common—*one or more short waves occur in the movements of imports which need not occur in those of consumption.* This seems to be the interesting result our analysis leads to. The result appears to be in agreement with the facts. A glance at Figures 1–5 shows the tendency to two-year and three-year cycles in imports which do not, or not with the same amplitude, occur in consumption.

V. COMPARISON WITH OTHER TYPES OF CYCLES

Quite a number of different cycles are already known to the student of business cycles.

Apart from seasonal movements, which need not occupy us here, we may distinguish between general cycles and specific cycles. Although the difference is not so sharp at closer examination as it would seem at first sight, general cycles occur in all, or almost all, fields of economic activity, whereas specific cycles are more or less restricted to special branches. The general cycles, for which Schumpeter recently proposed the names "Kondratiefs," "Juglars," and "Kitchins," show periods of some forty, eight, and three years, respectively. It is a well-known fact that these figures are subject to wide margins of uncertainty.

Specific cycles are now known in a wide variety. A coffee cycle of about fifteen years, a housing cycle in the United States[3] of about the same period, and a pig cycle of about three years have been studied.

Apart from these there is one type of cycle which is neither general nor specific in that it is common to a number of agricultural markets, viz., the two-year crop cycle. This cycle, to which ample reference has been made by Dr. J. Åkerman[4] and which was also emphasized, in the special case of sugar, by Henry Schultz,[5] is caused by the co-operation of the following two relations: (*a*) the crop of an agricultural product determines its price, by virture of the demand function, and (*b*) its price determines the acreage, and therefore to a certain extent the crop, of the following year. This cycle is very much disturbed by the big random fluctuations in yield per acre, by which the relation between acreage and crop is loosened, but nevertheless many traces of it can be found.

The question may even be put whether or not the two-year cycles which we have attributed to the acceleration principle for stocks are in reality only examples of the crop cycles. Since the characteristic of the latter is the opposite movements of prices and quantities, it is easy to test this assumption by comparing our figures with price figures. It appears that the two-year fluctuations in imports are, in general, not inversely correlated with similar two-year cycles in prices. Therefore, the short cycles now under discussion are of a different type.

[3] I tried to explain this cycle in my study, *Business Cycles in the United States* ("Statistical Testing of Business-Cycle Theories," Vol. II [Geneva: League of Nations, 1938]).

[4] *Om det ekonomiska livets rytmik* (Stockholm, 1928).

[5] *Statistical Laws of Demand and Supply with Special Application to Sugar* (Chicago, 1928).

VI. A SIGNIFICANCE PROBLEM IN THE STATISTICAL DETERMINATION OF α

The statistical determination of the coefficient α in equation (3′) offers a difficulty. Writing it in the form

$$u_t - u_t' = \alpha(u_t - u_{t-1}) \; , \tag{3''}$$

we see at once that, if u_t is subject to some error of measurement, this error will appear at both sides of the equation and therefore lead to spurious correlation. We shall try to find the influence on α by using a method similar to Frisch's[6], that is, by assuming that each of the variables consists of a systematic part and an error, whereas these errors are supposed to be uncorrelated with the systematic parts of any variable. We shall simplify the problem by assuming that u_t' is not subject to error—in fact, these errors would do no special harm.

Writing u_t'' for $u_t - u_t'$, Δu_t for $u_t - u_{t-1}$; v_t'' for the systematic part in $u_t'' = u_t - u_t'$; w_t for the error in $u_t - u_t'$; Δv_t for the systematic part in Δu_t ; and Δw_t for the error in Δu_t (which is indeed, under our hypothesis, equal to $w_t - w_{t-1}$), we have, when omitting the suffix t,

$$u'' = v'' + w \; , \tag{8}$$

$$\Delta u = \Delta v + \Delta w \; . \tag{9}$$

Our hypotheses are that

$$\Sigma v'' w = 0 \; , \tag{10}$$

$$\Sigma \Delta v \Delta w = 0 \; , \tag{11}$$

$$\Sigma v'' \Delta w = 0 \; , \tag{12}$$

$$\Sigma (\Delta v) w = 0 \; . \tag{13}$$

The difference from the problem treated by Frisch is that in our present problem the errors of the two variables depend on each other in a very definite way, whereas Frisch supposed them to be uncorrelated. We shall assume, however, that the error series w shows serial correlation zero:

$$\Sigma w w_{-1} = 0 \; .$$

It follows that

$$\Sigma w \Delta w = \Sigma w (w - w_{-1}) = \Sigma w^2 \; , \tag{14}$$

and

[6] R. Frisch, *Statistical Confluence Analysis by Means of Complete Regression Systems* ("Oslo Universitetets Økon. Inst.," No. 5 [Oslo, 1934]), p. 51.

$$\Sigma(\Delta w)^2 = \Sigma(w - w_{-1})^2 = 2\Sigma w^2 \,. \tag{15}$$

The last equation is only approximately fulfilled, but the relations (10)—(14) are of the same character.

Following Frisch, we call, in contradistinction to the "observed" regression coefficient $b = (\Sigma u''\Delta u)/(\Sigma\Delta u^2)$, the "true" regression coefficient a the one between v'' and Δv, and we assume the correlation between these variables to be perfect. Our object is to find a, being given the moments of the measurable variables u'' and Δu.

The assumption of perfect correlation between v'' and Δv translates itself into

$$a = \frac{\Sigma v''\Delta v}{\Sigma\Delta v^2} = \frac{\Sigma v''^2}{\Sigma v''\Delta v} \,. \tag{16}$$

Of these two expressions for a, one will yield us the value of Σw^2, whereas the other will yield us the value of a. The moments occurring in (16) may be reduced to measurable moments by the following transformations:

$$\Sigma u''^2 = \Sigma(v'' + w)^2 = \Sigma v''^2 + 2\Sigma v''w + \Sigma w^2 = \text{(using [10])}$$
$$= \Sigma v''^2 + \Sigma w^2.$$

$$\Sigma u''\Delta u = \Sigma(v'' + w)(\Delta v + \Delta w) = \Sigma v''\Delta v + \Sigma w\Delta v$$
$$+ \Sigma v''\Delta w + \Sigma w\Delta w = \text{(using [13] and [12])}$$
$$= \Sigma v''\Delta v + \Sigma w\Delta w = \text{(using [14])} = \sum v''\,\Delta v + \sum w^2.$$

$$\Sigma\Delta u^2 = \Sigma(\Delta v + \Delta w)^2 = \Sigma\Delta v^2 + 2\Sigma\Delta v\Delta w + \Sigma\Delta w^2 = \text{(using [11])}$$
$$= \Sigma\Delta v^2 + \Sigma\Delta w^2 = \text{(using [15])} = \Sigma\Delta v^2 + 2\Sigma w^2 \,.$$

It follows that

$$\Sigma v''^2 = \Sigma u''^2 - \Sigma w^2 \,, \tag{17}$$

$$\Sigma v''\Delta v = \Sigma u''\Delta u - \Sigma w^2 \,, \tag{18}$$

$$\Sigma\Delta v^2 = \Sigma\Delta u^2 - 2\Sigma w^2 \,. \tag{19}$$

Indeed we find that only measurable moments and Σw^2 enter into these expressions. The simplest procedure is to find Σw^2 and a by numerical calculation in each case. This may be demonstrated by the example of cotton imports. For this purpose we have subdivided the period on which our figures bear—1854–1912—into three periods (corresponding to the ups and downs of the "Kondratiefs"[7]: (I) 1854–73, (II) 1874–95, and (III) 1896–1912.

[7] Following Schumpeter, I use this name to indicate the so-called long waves.

TABLE 2

Period	$\Sigma u''^2$	$\Sigma u''\Delta u$	$\Sigma \Delta u^2$	Σw^2	$b=\frac{\Sigma u''\Delta u}{\Sigma\Delta u^2}$	$a=\frac{\Sigma u''\Delta u-\Sigma w^2}{\Sigma\Delta u^2-2\Sigma w^2}$
I ...	1149	2402	8875	780	0.27	0.22
II ...	1784	2146	6492	1730	0.33	0.13
III ...	4702	6034	12005	340	0.50	0.50

The data and results are given in Table 2. According to these results, the observed regression coefficients would not much be overvalued by spurious correlation in Periods I and III, but it would in Period II. In this case, however, we find that Σw^2 is almost as large as $\Sigma u''^2$. This is distinctly too pessimistic as to the reliability of the material; and it would seem more probable, therefore, that our statistical theory is too rigorous. Especially the assumption that there is perfect correlation between v'' and Δv alone may be too simple. Other factors, too, may determine stocking demand. Therefore, a slightly different approach may also be tried.

All our hypotheses may be maintained with the exception of perfect correlation between v'' and Δv. Instead it may be assumed that the errors in u'' are at a maximum such that their variance is half that of u'' itself, that is, $\Sigma w^2 \leqq \frac{1}{2}\Sigma u''^2$. Since now we assume that v'' may depend on other factors as Δv only, the first regression $(\Sigma v''\Delta v)/(\Sigma \Delta v^2)$ should be taken. This equals

$$\frac{\Sigma u''\Delta u - \Sigma w^2}{\Sigma \Delta u^2 - 2\Sigma w^2} .$$

As long as

$$\frac{\Sigma u''\Delta u}{\Sigma \Delta u^2} < \frac{1}{2} ,$$

as in I and II, the minimum value will be obtained for Σw^2 at its maximum value; and this turns out to be 0.24 for I, 0.22 for II, and 0.50 for III. In this light, even for Period II the effect of spurious correlation does not seem to be serious. It will easily be seen that the observed regression coefficient will always be correct, regardless of the extent of the errors, if it is 0.50, as in III.

DYNAMICS OF COMMODITY PRICES

CHARLES F. ROOS
Institute of Applied Econometrics

I. INTRODUCTION

CHANGES in commodity prices are of major importance to any national economy and have long held important places in economic theory. Kuznets has made estimates of the extent to which savings (as reflected in the usual national income statements) reflect inventory profits and losses and reached the conclusion that inventory gains in the United States in 1919 were over two billion dollars; in 1921, only two years later, the inventory loss was over four billion dollars; in 1924 the gain was over two billion; in 1930 the loss, over four billion; and in 1933 the gain, over two billion.[1] Inventory losses and gains have thus amounted to $1\frac{1}{2}$ to 10 per cent of the national income.

Arthur has studied the relation of inventory profit to business cycles and reached the conclusion:

> The greatest portion of industrial inventories is analagous to water in the pipes of the economic production system. They cannot be drawn off and consumed without stopping the operations of the system. An increase in the prices at which these inventories are valued produces an apparent profit, but this gain is fictitious and unexpendable; it could not be converted into cash without liquidating the inventories. Such a profit is in effect an unrealized (and in practice largely unrealizable) capital gain, yet it has been treated by most accountants, statisticians, and economists as current income—even as a part of our national income.
>
> It is a comparatively easy matter to eliminate unrealized inventory gains and losses from statements of business income, but the methods for accomplishing this have not received any general acceptance. Businessmen make their decisions as though fictitious inventory gains (amounting in some years to several billions of dollars) were expendable, and the effect of this misunderstanding is to aggravate the cyclical fluctuations of business. The elimination of fictitious inventory gains and losses from current income statements would be an important contribution toward moderating cyclical extremes.[2]

The main factors influencing the price of a particular commodity —cost of production, demand, old and new supply, price changes, and

[1] Simon Kuznets, *Commodity Flow and Capital Formation*, p. 408.
[2] Henry Arthur, "Inventory Profit in the Business Cycle," *Journal of the American Statistical Association*, 1936.

monetary changes—have long been known. Their effects have also been rather well understood, at least in a qualitative way.[3] But still timely are Irving Fisher's remarks of 1892:

> Water seeks its level, but this law does not fully explain Niagara. A great deal of special data are here necessary and the physicist is as unfit to advise the captain of the *Maid o' the Mist* as an economist to direct a Wall Street speculator. The failure to separate statics from dynamics appears historically to explain the great confusion in early physical ideas. To make this separation required the reluctant transition from the actual world to the ideal. The actual world both physical and economic has no equilibrium. . . . . The dynamical side of economics has never yet received systematic treatment. When it has, it will reconcile much of the present apparent contradiction, e.g., if a market is out of equilibrium things may sell for "more than they are worth," as every practical man knows: that is the proper ratios of marginal utilities and prices are not preserved.[4]

The present paper which is devoted to a study of the dynamics of commodity prices will draw upon the theories and methods developed by that great pioneer, Henry Schultz,[5] whose untimely death has left such a gap to be filled in research personnel. Subjects investigated include the relative importance of national income, general price level, monetary factors and speculative demand, in influencing commodity price movements,[6] and the extent to which major price changes can be anticipated.

For the modern functional treatment of demand or price we can do no better than refer to the comprehensive work, *The Theory and Measurement of Demand,* by Henry Schultz, who spent ten years studying the problem under a grant of the Social Science Research Council. In recognizing the need for a theory of demand more general than that existing at the time, Professor Schultz ably stated the modern concept:

[3] W. S. Jevons believed there was a sequence as follows: "Cost of production determines supply, Supply determines final degree of utility, Final degree of utility determines value" (*Political Economy* [London, 1871], p. 165). See also Léon Walras, *Eléments d'économie politique pure* (Paris, 1874); Auspitz and Lieben, *Untersuch ungen über die Theorie des Preises* (Leipzig, 1889); Augustin Cournot, *Theorie des richesses* (Paris, 1838); and E. von Böhm-Bawerk, "Wert, Kosten und Grenzmutzen," *Jahrbucher für Nationökonomie und Statistik,* III, No. 3, 321–78.

Irving Fisher, in his *Mathematical Investigations in the Theory of Value and Prices* (1892), gave the first systematic treatment of the problem. His work contains a bibliography of early work on the subject.

[4] *Op. cit.,* p. 104.

[5] See, especially, Henry Schultz, *Statistical Laws of Demand and Supply* (Chicago, 1928); "The Shifting Demand for Agricultural Commodities, 1875–1929," *Journal of Farm Economics,* XIV (1932), 201–27; *The Theory and Measurement of Demand* (Chicago, 1938).

[6] See, especially, C. F. Roos, "A Dynamical Theory of Economics," *Journal of Political Economy,* XXXV (1927), 632–56; "A Mathematical Theory of Price and Production Fluctuations and Economic Crises," *ibid.,* XXXVIII (1930), 501–22; and *Dynamic Economics* (Bloomington, 1934).

The law of demand of the mathematical school

$$D = f(p_1, p_2, \cdots, p_n)$$

includes the Cournot-Marshall law of demand,

$$D = f(p)$$

as a special case. To obtain the latter from the former we make use of the only valid non-experimental method for keeping the "other things" constant which may be stated as follows: First, take all the factors (variables) into consideration; second, assign constant values to all variables except the price and the quantity of the commodity in question. This means that we must determine the equation connecting the quantity demanded and all the prices "and all other important other variables" and then assign constant values to all the variables except the two under consideration.

The designation of the ordinary demand curve as a special case of the general demand function, from which it is derived by assigning constant values to all the variables except the price and quantity under consideration, marks a distinct improvement over the classical and neo-classical conception of this curve. The neo-classical economists, though they talked about other variables, never took the pains to first introduce them into their demand equation and then to assign them constant values. These economists never thought to raise the question whether it was always possible to keep other things constant, nor did they ever face the problem of the levels at which each of the "other things" might be kept constant.[7]

II. THE GENERAL PRICE LEVEL

A customary procedure in relating price and demand or price and supply is to "deflate" price by an index of commodity prices.[8] This has usually been assumed to be equivalent to removing the effects of factors which affect all commodities, that is, national income, general speculation, foreign exchange, and the like. The "deflated" price is then related to consumption and a time trend, or consumption, national income and a time trend, or to supply and these factors.

Study of commodity prices shows, however, that changes in the

[7] In November, 1927, the author first discussed the new concept of demand which Schultz immediately reviewed and later developed. In his first review Schultz said: "C. F. Roos, National Research Fellow in Mathematics, introduces into his demand and cost functions not only the quantities produced and the prices paid, but also the time element, and the rates of change of production, prices, with respect to time; and he discusses the phenomena of competition, monopoly and cooperation, for these general functions. By introducing the rate-of-change concept he has made his problem one in economic dynamics, not in economic statics. . . . The mathematical theory of economics has recently been changed from a static to a dynamic theory, and has been stated in a form admitting of quantitative application. It therefore is both a challenge and an invitation to all economists to verify it or to improve upon it." See also C. F. Roos and Victor S. von Szeliski, "The Concept of Demand and Price Elasticity," *Journal of the American Statistical Association*, XXXIV (1939), 653–59.

[8] Such is the procedure used by such writers as Henry Schultz, Holbrook Working, Louis Bean, Mordecai Ezekiel, and others.

deflator or general factor account for the major portion of the variations in price; that is, that from the point of view of understanding the dynamical side of prices, very little has been uncovered. The major statistical discoveries that have been made are due to economists of the United States Department of Agriculture. Mordecai Ezekiel and others in the Department of Agriculture as early as 1928 were drawing a free-hand curve to approximate a relationship between two variables such as cotton consumption and prices of cotton, plotting the residuals (actual consumption minus consumption represented by the free-hand curve) against a third variable such as *industrial activity,* and drawing a free-hand curve to represent a relationship between the residuals described above and industrial activity.[9] Thus, their view was that the same factors that affect industrial activity affect commodity prices. And Irving Fisher, in a paper before the American Association for the Advancement of Science in December, 1932, developed his former observation that changes in business activity tend to be related to changes in the level of commodity prices.[10] Subsequently, 1933 to date, Carvel Lange has used this premise to forecast price changes by a method of graphical harmonic analysis applied to the Moody business index. With this approach Lange has been able to predict with better than chance accuracy changes in the Moody index and in commodity prices.

The author, while with the National Recovery Administration in 1934, observed that manufacturers tended to buy commodities in line with their current new orders and hence that fluctuations in these new orders, whatever their cause, must have important effects on commodity prices. The problem of determining general price movements seemed, therefore, to be practically identical with that of determining the economic and psychological forces which cause changes in the volume of new orders. The first problem, however, was that of constructing new order indexes since none existed and data for making them were scarce. The next problem was that of studying the relation of the new indexes to commodity price indexes.

On Chart I is shown an index of new orders received by manufacturers of cotton, rayon, wool, and silk textiles, and clothing, leather goods, copper and brass goods, electrical appliances, furniture, carpets and rugs, house furnishings, automobiles and parts, building supplies, refrigerators, washing machines, machinery, machine tools, foundry equipment, cranes, electrical equipment, electric motors,

[9] See Ezekiel, *Methods of Correlation Analysis* (New York, 1930), or Roos, *Dynamic Economics*, pp. 45–52.

[10] C. F. Roos (ed.), *Stabilization of Employment* (Bloomington, 1933), chap. x.

CHART I
New Orders Received by Manufacturers and Production

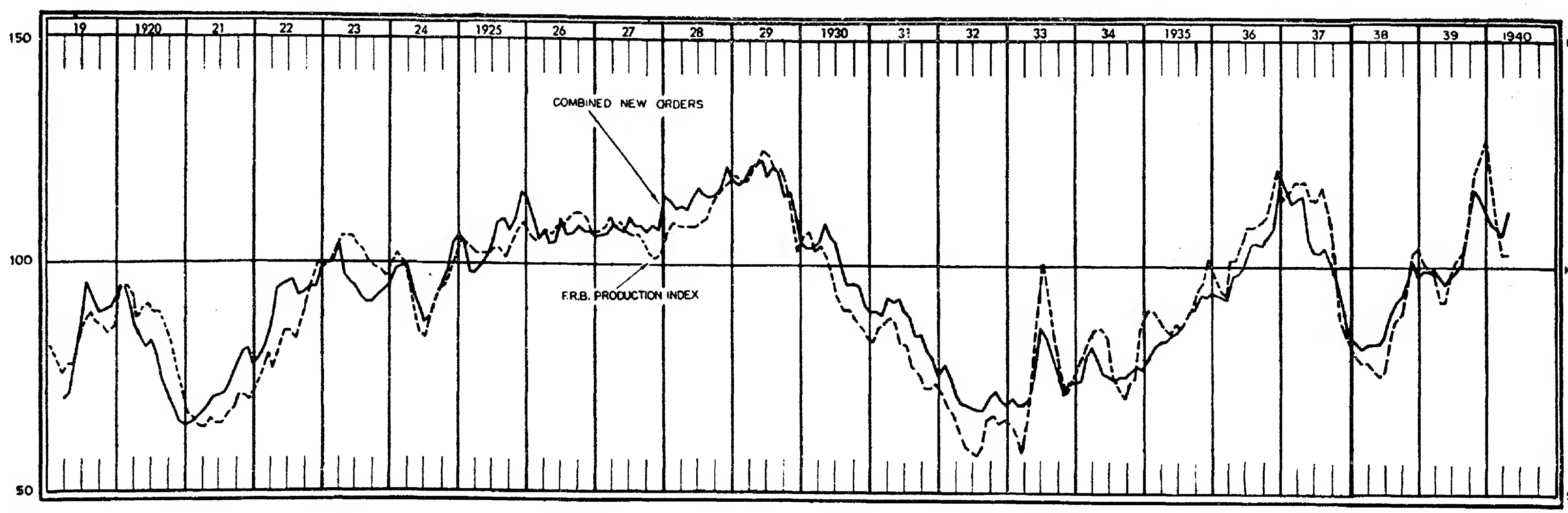
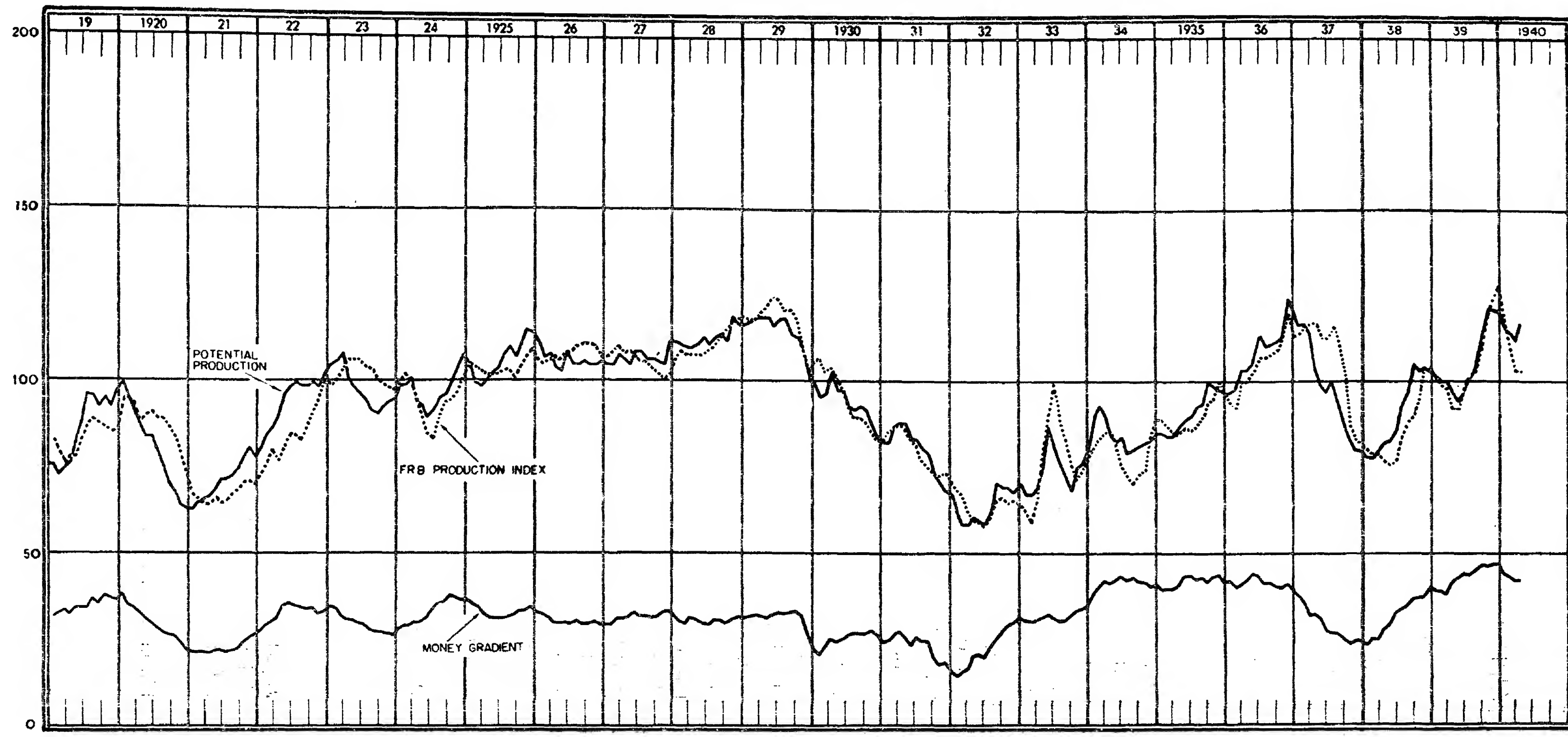


CHART II

Relation of New Orders and Monetary Change to Production



woodworking machinery, machine accessories, steel, shipbuilding, valves and fittings, etc. In this index of recent years the physical volume of orders of over nine hundred companies in the above industries, seasonally corrected, have been used; in earlier years, as few as a hundred companies. The weights taken by the industries are based upon their importance as shown by the United States Census of Manufactures. The index base is the same as that of the Federal Reserve Board's index of industrial production, which is also shown on this chart.

Study of the chart shows that there are several important differences in the contours of the two curves, notably in 1934, when the Board used a poor seasonal because of the change in introduction dates of new automobile models. They also differ because the order curve tends to lead production, especially at the major tops and bottoms.

An important measure of likely changes in orders and production is the rate of change of demand deposits, or money gradient,[11] as such a series is called on Chart II. This is because individual manufacturers watch closely the changes in their cash assets and tend to gauge their production and other operations by these changes, increasing production when the rate of change of their own deposits has attained and maintained a higher level and decreasing them after this rate has fallen. If deposits are expanding, most manufacturers are experiencing increased deposits and hence planning increased production or expenditures for equipment. Also shown on Chart II is a weighted average of the money gradient and the new orders index of Chart I, which is called "potential production."

The index "potential production" is a measure of the normal demand for industrial commodities. From the classical theory we would therefore expect a relationship between it, an index of stocks of industrial commodities, and an industrial price index of the form

$$\text{Price} = (\text{Potential Demand})^{\alpha}\ (\text{Stocks})^{\beta}\ \text{Constant}\,,$$

where α is the elasticity of demand and β the elasticity of supply. In mathematical symbols this becomes

$$\pi(x) = AD(x)^{\alpha}\,S(x)^{\beta},$$

where $\pi(x)$ represents the price index at the time x, $D(x)$ is the demand at time x, $S(x)$ represents stocks, and A is a constant.

On Chart III are shown the Irving Fisher index of industrial

[11] This is a ten months' moving average, with nearest months weighted most.

prices and an index of visible stocks of such commodities as iron ore, zinc, lead, copper, cotton, rubber, and lumber, constructed from data supplied by the Department of Commerce. If this index of visible stocks and the potential demand of Chart II are used in the right-hand side of the above formula, a reasonably good fit to the major trend of the Fisher index of industrial prices can be obtained.

But such formulation obviously takes no account of speculative demand, of changes in foreign demand as represented by changes in the exchange rates, or of institutional developments such as can be represented by a time trend. A better representation of price $\pi(x)$ would be given by

$$\pi(x) = AD(x)^{\alpha_1} S(x)^{\beta_1}, E(x)^{\gamma_1}$$

where

$$E(x) = \frac{r(x-1) + \Delta r}{r(x-1)} = \frac{r(x)}{r(x-1)} .$$

Δr is the change in a foreign exchange index $r(x)$ shown on Chart III and $r(x-1)$ is the value of this index the month before. The rate of change of exchange is preferable to the exchange rate itself, because after a short time the effects of new levels of exchange are reflected in new orders and hence taken into account by another part of the formula. The average time lag of this effect, for the Fisher index of industrial prices, is three months.

There still remains the problem of accounting for purely speculative price changes. These often depend on past price moves and past levels of orders. This lag effect can be explained by using a formula in which the price factor appears as follows:

$$P(x+n) = A[\pi_1(x)]^{\alpha_2}\left[1 + \frac{\Delta P}{\int_{x-a_0}^{x-1} P(t)\,dt}\right]^{\beta_2} [P(x)]^{\beta_3},$$

where

$$\Delta P = P(x) - P(a_0).$$

This formulation says merely that the present price level depends upon the percentage change in level in the recent past.

The function

$$1 + \frac{\Delta P}{\int_{x-a_0}^{x-1} P(t)\,dt}$$

or

$$1 + \frac{P(x) - P(a_0)}{\sum_{x-a_0}^{x} P(x)} ,$$

which is useful in studies of commodity price movement, might well be called the *speculative carry-over*.[12] It can be derived as follows:

$$1+\frac{\Delta P}{\sum\limits_{x-a_0}^{x} P(x)}=1+\frac{P(x)-P(x-a_0)}{\sum\limits_{x-a_0}^{x} P(x)}=\frac{\sum\limits_{x-a_0+1}^{x} P(x)}{\sum\limits_{x-a_0}^{x-1} P(x)}=\frac{\bar{P}(x)}{\bar{P}(x-1)}\,.$$

A correlation analysis shows that $a_0 =$ six months. Past demand and past stocks can best be taken into account by using the integral of $\pi(x)$ over the preceding five months, that is, $\pi = \int_{x-4}^{x} \pi(t)\,dt$ or its approximation $\sum\limits_{x-4}^{x} \pi(x)$.

There still remains a time trend factor in the residuals. The trend obtained gradually slopes downward except for 1919–20, where the decline was steeper. Combining all these factors, one obtains

$$P(x+2)=A\left[\frac{\bar{P}(x)}{\bar{P}(x-1)}\right]^{-\frac{1}{2}}[P(x)]^{.70}\,[\pi(x)]^{1.08}\,S^{.92}\,T(x)$$

where $T(x)$ is a time trend.

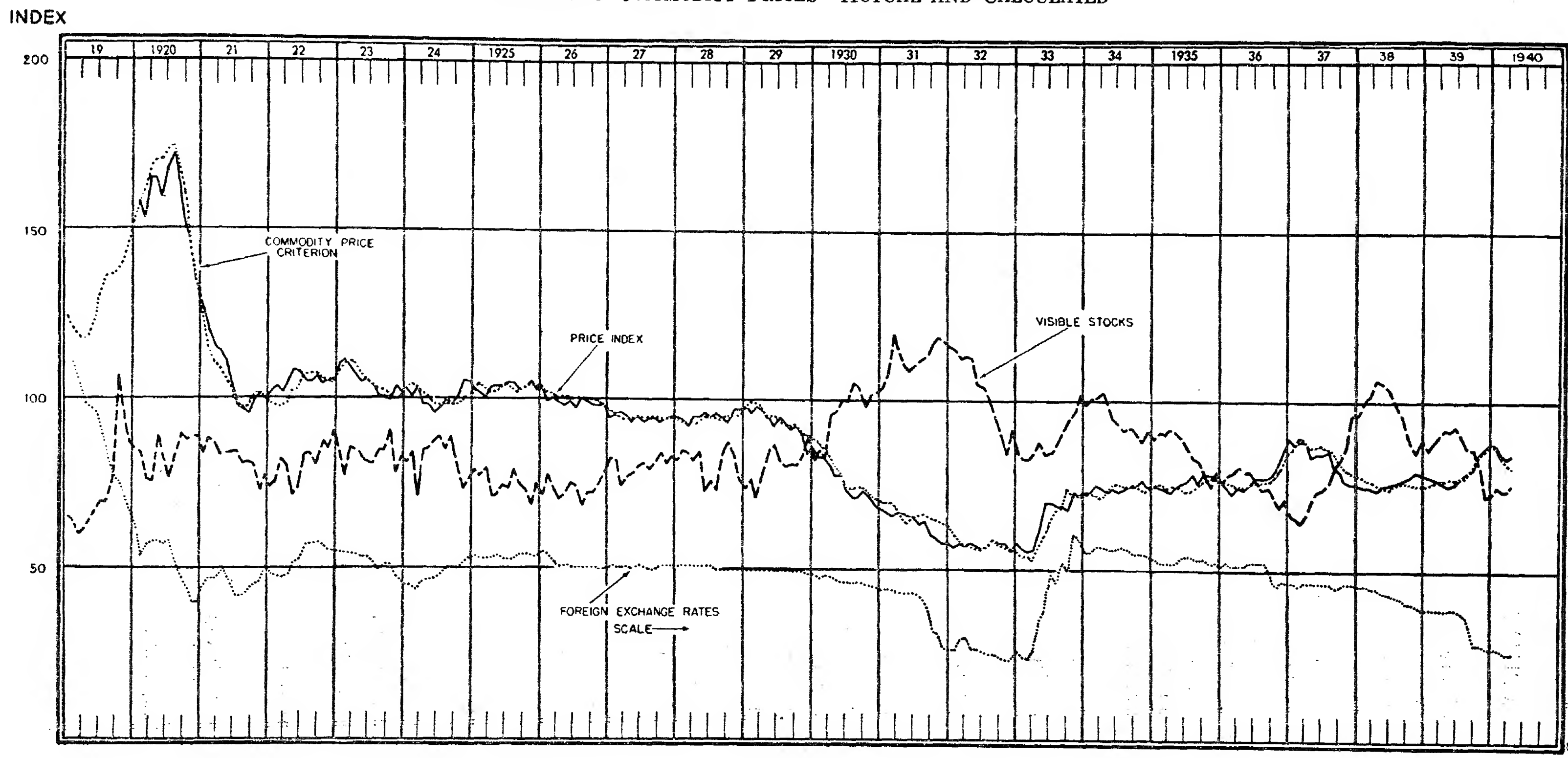
A curve derived from a formula which fits the Irving Fisher industrial price index is shown on Chart III with this price index. There are twelve constants involved in the formula, counting the lag constants and the time trend. The lag constants were determined empirically. Only six constants, exclusive of the three in the time trend, were determined by the method of least squares. Consequently, the fit is remarkably good for the entire period monthly from 1920 to 1938, that is, for the 228 observations from which the constants were obtained and also for 1939 and 1940, which represent extrapolations.[13]

An interesting point is that the calculated series has a persistent lead of about two months over the actual price series. Hence, here is a genuine forecasting series, at least until it becomes too widely used.

[12] The function "speculative carry-over" involves the notion of *economic* hysteresis, which the author introduced into economic literature (see C. F. Roos, "A Mathematical Theory of Competition," *American Journal of Mathematics*, XLVII [1925], 163–75).

[13] How to compute the standard error of regression coefficients derived from time series is a matter of controversy. The author holds the view of Walter Shewhart, Corrado Gini, and others that the errors formulas of sampling theory apply only to cases in which samples are drawn from constant universes and that, in general, economic time series do not satisfy this condition. Also there is the important matter of choice of function, which is not covered by the usual theory (see, e.g., Roos, *Dynamic Economics*, pp. 246–54, and Roos, "Annual Survey of Statistical Techniques," *Econometrica*, October, 1935).

CHART III
Industrial Commodity Prices—Actual and Calculated



Particularly interesting periods are January-December, 1929, April, May, and June, 1933, and January, 1937, to the present time. In each of these periods of sharp price moves the calculated series gave early and correct warnings of changes in major trend.

Note in particular that the cost of production, which plays a central role in economic theory, has been given no direct place in the theory of changes in the price level here presented. It is, however, given indirect attention through the factor of stocks, which is related to previous excess of price over cost of production. Also, the term "speculative carry-over" indirectly accounts for changes in costs of production resulting from previous changes in the general price level. In general, cost of production does not have to be used to explain price or price change of raw materials. However, in some cases of individual price studies, such as copper, cost of production is an important factor.

III. SOME OBSERVATIONS ON CAUSES OF CHANGES IN THE PRICE LEVEL

It has just been shown that price level as a function of time can be represented by the level of new orders received by business, commodity stocks, the rate of change of bank deposits, the rate of change of foreign exchange, and speculative carry-over. Whether this represents an "explanation" of price levels depends upon the extent of one's curiosity.[14] One can, if he wishes, insist upon explanations of each of the above component parts and then in turn explanations of each new component almost *ad infinitum*. The usual scheme is to carry the explanation back to utility or human desires or caprices and let the matter rest there.

In the present case, as is true in most economic situations, the factors react on one another to a considerable degree. For instance, changes in commercial loans, a substantial part of bank deposits, result from previous changes in new orders and business working capital; the last, in turn, is highly correlated with previous bank deposits. To continue: speculative excesses appear to develop as a result of previous maladjustments between new orders received by business and retail consumption.

A broad index of agricultural prices such as the Fisher index can likewise be explained by the same approach. Stocks of agricultural commodities and plantings or expected stocks replace industrial stocks,

[14] P. Bridgman (*The Logic of Modern Physics* [New York, 1927], p. 37) says: "The essence of an explanation consists in reducing a situation to elements with which we are so familiar that we accept them as a matter of course, and our curiosity rests."

and the other factors take slightly different weights. But otherwise the analysis is the same.

IV. STUDIES OF PRICES OF INDIVIDUAL COMMODITIES

From the above discussion it is clear that studies of prices which proceed along the classical lines of deflating for general price level and then correlating with specific factors affecting the commodity can be materially improved by using a theoretical "deflator" such as the commodity price criterion of Section II. This has the obvious advantage that it permits one to understand the effects of the factors that determine the commodity price level—the general demand, the general supply, monetary changes, foreign exchange, and *speculative carry-over.* These important factors in the approach given here are measured separately and are not simply combined into a general index, which measures their ultimate or *later* effect.

It is perhaps worth adding that, where commodity price studies are to be used for business or government policy-making decisions, it is imperative that a general "explanation" or criterion rather than a deflating index be used because by this procedure a lead is given over actual prices, and this lead may be the sole difference between the reaching of a successful or an unsuccessful policy decision. Moreover, a study of Charts I, II, and III in conjunction with the price of any commodity will show that the really important price changes are induced by the general economic forces rather than by specific factors affecting the commodity. Hence one may say that unless the dynamics of commodity price movements as a whole are understood and the important factors are successfully anticipated, the chances of successful speculation in commodities are small. Since this last is extremely difficult, it is no wonder, therefore, that business has experienced the great speculative losses and gains indicated by Kuznets and cited in the introduction (Sec. I).

Our studies indicate that the soundest approach to the price of an individual commodity is to study its relation to the specific factors and to such general factors as (1) national income or, better, if a lead over prices is desired, expected income, calculated from an index of new orders, government spending and agricultural marketings; (2) changes in the money supply or money gradient; and (3) changes in foreign exchange or, if a lead is desired, expected changes as measured by relative price levels,[15] trade balances, and capital move-

[15] See, e.g., Gustav Cassel, "Quantitative Thinking in Economics."

CHART IV-A

COPPER PRICES—ACTUAL AND CALCULATED

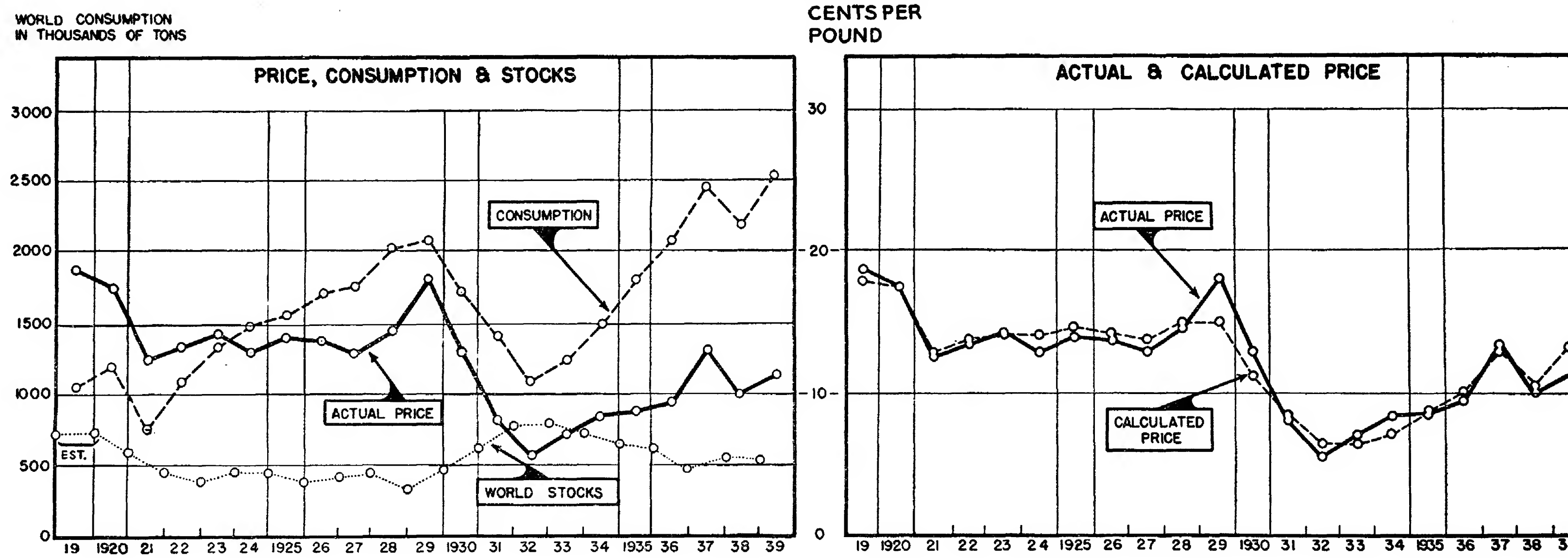
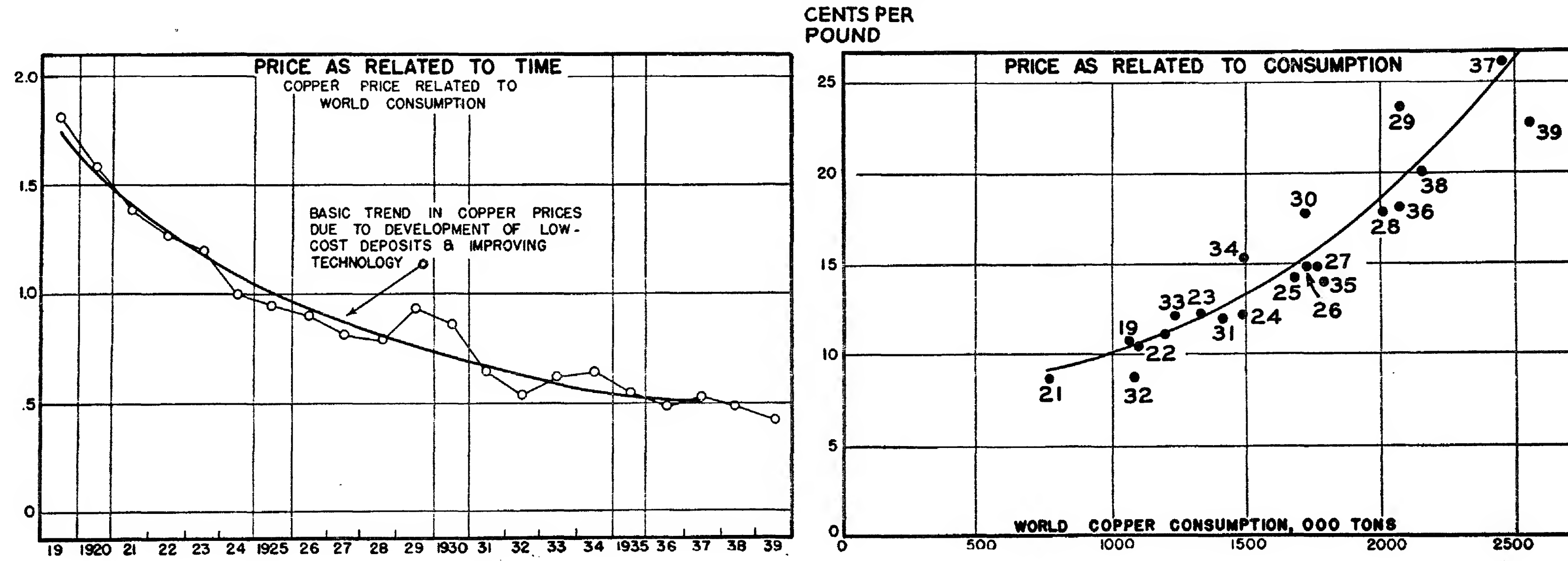


CHART IV-B
COPPER PRICES—ACTUAL AND CALCULATED



ments. The reason is that the weights assumed by the factors (1), (2), and (3) are not the same for different commodities. Hide prices, for instance, have an elasticity of one for changes in foreign exchange, whereas the elasticity for plaster prices appears to be zero.

V. SOME EXAMPLES—COPPER, WOOL, AND HIDES

A. COPPER PRICES

The United States price of copper is interesting as an exception in that cost of production plays a major role, and major factors which are important for the general price level have only minor influences.[16]

Copper prices in the United States can be rather well explained in terms of only world-consumption and a time trend which measures lowered cost of production due to the development of low-cost deposits and improving technology. This is shown on Chart IV. The fit is extraordinarily good, with principal residuals being in 1929, 1930, 1932, and 1934. The curve in the lower right-hand panel comes rather close to representing a supply curve for copper in the static sense.

An examination of the lower left-hand panel of Chart IV reveals that the deviations have a systematic pattern and hence that the fit in these and other years can be materially improved by introducing a term to represent speculative carry-over. In fact, world-consumption, cost of production, and speculative carry-over yield a rather complete explanation of copper prices.

Several factors which might be thought to have an important influence on copper price can be omitted from formulas explaining price. World-stocks, for instance, can be ignored because stocks have tended to be inversely correlated with consumption, as is shown by the upper left-hand panel. It is not necessary to introduce the general demand factor unless one wishes to forecast copper prices.[17] The foreign exchange gradient can be ignored because United States copper prices are protected by a tariff which tends to insulate them to a certain extent from world-currency changes. Examination of the residuals, however, indicates that a slight improvement in fit could be obtained by introducing this gradient or the rate of change of foreign

[16] The following study is taken from the author's paper presented to the Econometric Society, at their Atlantic City meeting, December, 1937. The values for 1938 and 1939 are extrapolations.

[17] The Institute of Applied Econometrics has studies of the consumption of copper by industries which enables it to estimate accurately this important demand factor many months ahead. United States consumption is about one-third of world-consumption and is correlated with it but fluctuates more violently. United States consumption is correlated with the Institute's index of new orders received by business because copper is used in most industries.

exchange. This is probably because the United States exports a small amount of copper.

It is thus evident that the problem of determining the dynamics of copper prices can be formulated in several different satisfactory ways. The reason is that various economic series which can be used in the explanations are highly correlated.

B. WOOL PRICES

The domestic supply of wool, unlike that of copper, is far from adequate to supply domestic needs. Hence wool prices are quite sensitive to changes in foreign exchange, particularly to changes in British and South American currencies.

The price of fine territory wool (63s70s80s as reported by the Bureau of Agricultural Economics) can be very well represented by a formula of the type

$$\text{Price} = \begin{pmatrix}\text{Consumer}\\ \text{Income}\end{pmatrix}^{\alpha} \begin{pmatrix}\text{Domestic}\\ \text{Production}\end{pmatrix}^{\beta} \begin{pmatrix}\text{Foreign}\\ \text{Production}\end{pmatrix}^{\gamma} \begin{pmatrix}\text{Domestic}\\ \text{plus}\\ \text{Foreign}\\ \text{Stocks}\end{pmatrix}^{\delta} \begin{pmatrix}\text{Foreign}\\ \text{Exchange}\\ \text{Gradient}\end{pmatrix}^{\eta} \begin{matrix}\text{Time}\\ \text{Trend}\end{matrix}.$$

The most appropriate measure of consumer income, in this case, is lower bracket income.[18] Its independent effect (other factors held constant) on wool prices is shown in Table 1.

TABLE 1

WOOL PRICES AND THE INCOME FACTOR

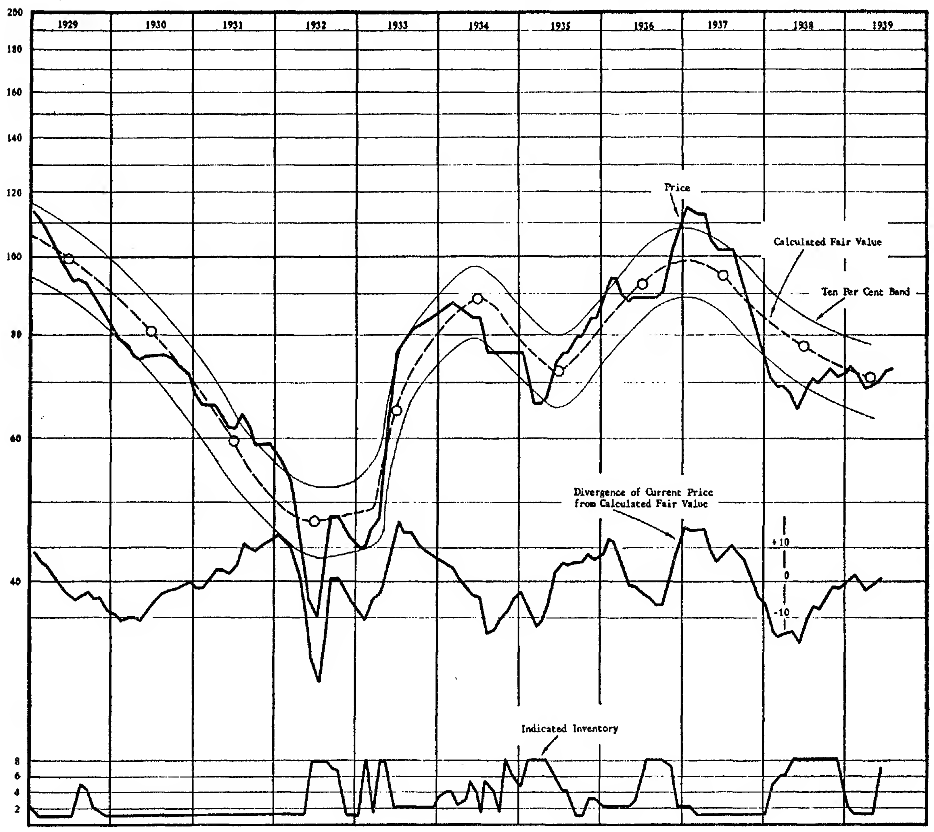
Income Billions	Wool Prices Cents per Pound
35	67.0
40	75.2
45	83.3
50	91.4
55	99.5
60	107.5
65	115.3

With other factors at their average values of the last fifteen years, a 35-billion-dollar income would support a 67-cent wool price and a 65-billion-dollar income a 115-cent price. The elasticity, α, is thus .88, which means that a 1 per cent increase in income causes an 0.88 per cent increase in price, other factors being unchanged.

[18] For a definition of lower bracket income see C. F. Roos and Victor S. von Szeliski, *Dynamics of Automobile Demand* (Detroit: General Motors Corporation, 1939), pp. 39–40.

Statistical analysis yields the following weights for the other factors. Each million pounds of domestic production of wool tends to lower the price 0.48 per cent; each million pounds of foreign production, 0.03 per cent. The apparent effect of increasing the beginning-of-year stocks by one million pounds is to decrease the price 0.04 per cent. The foreign exchange gradient has a powerful short-time effect on wool prices, the experience of recent years indicating that a 1 per cent

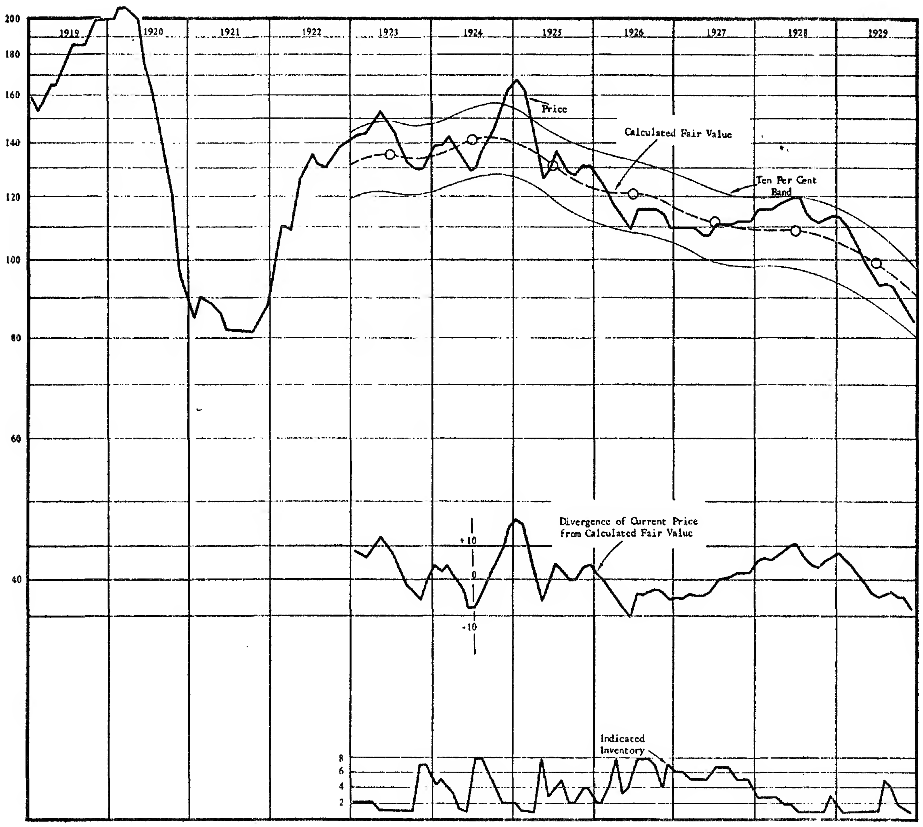
CHART V-A
WOOL PRICES—ACTUAL AND CALCULATED



increase in the former is accompanied by a 1.38 per cent increase in the latter. Domestic production of rayon staple is increasing in importance and accounts for a good part of the time trend.

Actual and calculated prices are shown on Chart V. Also shown on this chart is a "speculative band" within which the price tends to remain. Once more it is clear that most of the deviations between

CHART V-B
WOOL PRICES—ACTUAL AND CALCULATED



actual and calculated prices can be accounted for by a term representing "speculative carry-over." As Fisher remarked years ago:

> It may justly be claimed that not only do prices vary from day to day, but even if they did not, the individual's estimate of utility is fitful and, although at the instant he closes a bargain, his estimate of utility must be regarded as corresponding to the given price, yet he is likely generally and certain sometimes to regret his action so that if he were to live the year over again he would act very differently. . . . . First though prices vary from hour to hour under the influence of excitement and changing rumors, and from season to season under causes meteorological and otherwise, yet these fluctuations are self-corrective. . . . . Secondly the individual caprice is self-corrective. If a man lays in too large a stock of provisions this week, he will buy less next. The theory of probabilities therefore substantially harmonizes the theoretical and the actual.[19]

[19] *Op. cit.*, p. 21.

C. HIDE PRICES

In each of the above examples cost of production plays an important role—either directly, as in the case of copper prices, or indirectly and to a lesser extent through its influence on production, as in the case of wool.[20] We consider next an example in which the cost of production plays a negligible role—hides, which are by-products of the slaughter industry. As might be expected in such cases, existing stocks of hides or leather are more important than new production or cost of production.

Hide prices are about as volatile as those of any commodity, and therefore challenging. Green salted packers' hides at Chicago sold as high as 60 cents a pound in 1920 and as low as 4 cents in 1932. Moreover, there is an important seasonal variation in hide prices which also makes them interesting. Prices tend to be high during the summer, when the quality of hides from slaughter is highest, and low during the winter months, when the quality of hides from slaughter is lowest. The seasonal correction is 93.3 in July and 107.3 in January.

Factors which affect hide prices most are United States and foreign stocks of hides, United States and foreign stocks of leather, United States and foreign production, production and stocks of calf and kip skins, consumer purchasing power, consumption of leather, changes in foreign exchange rates, and changes in bank deposits.

Net imports of hides in the United States customarily account for about 10 per cent of the local supply. In our studies, however, imports have not been found useful for explaining price changes—in hides or any other commodity. This is because imports change as a result of currency changes and hence are measured by the foreign exchange factor or because imports are an effect, not a cause. High prices attract large imports; low prices discourage them. Once the hides are in this country or in transit here, they become part of the stocks and they then affect price in the usual direction, higher stocks causing lower prices and vice versa.

Stocks of hides include raw hides, stock in process, and finished leather converted into a hide equivalent. They are visible stocks. Once leather is "consumed," the Tanners' Council, the source of information on stocks of hides and leather, loses track of it statistically. These stocks include leather in process, finished products, and leather stocks held by distributors. They can amount to half a year's leather consumption. They can best be measured by a trailing average of consumption, an eight months' trailing average in the present study.

[20] Wool and mutton are called "completing commodities," (see, e.g., John Stuart Mill, *Political Economy*, Book III, chap. xvi; W. S. Jevons, *op. cit.*, p. 197; and Fisher, *op. cit.*, p. 65).

Another important invisible is the consumers' stocks of shoes and other finished goods in use or usable, which probably also exceed a six months' supply. Data for estimating this factor have been collected for only three years. However, these invisible stocks can be calculated by cumulating leather consumption and shoe production, as is done in the present study.[21]

Consumer purchasing power is measured by lower bracket income in this case, as was done also for wool.

A statistical analysis shows that

$$\text{Price} = \begin{pmatrix}\text{Consumer}\\ \text{Purchasing}\\ \text{Power}\end{pmatrix}^{\alpha} \begin{pmatrix}\text{Consumption}\\ \text{of}\\ \text{Leather}\end{pmatrix}^{\beta} \begin{pmatrix}\text{Cattle}\\ \text{Slaughter}\end{pmatrix}^{\gamma} \begin{pmatrix}\text{Stocks of}\\ \text{Hides and}\\ \text{Leather}\end{pmatrix}^{\delta} \begin{pmatrix}\text{Foreign}\\ \text{Exchange}\\ \text{Gradient}\end{pmatrix}^{\eta} \begin{matrix}\text{Time}\\ \text{Trend}\,.\end{matrix}$$

The elasticity of consumer purchasing power is 1.5; that is, a 1 per cent change in consumer purchasing power is accompanied by a 1.5 per cent change in price. The elasticity of consumption of leather is .75, of cattle slaughter .60, of stocks of hides and leather .70, of foreign exchange gradient 1.1. The time trend and residuals are shown on Charts VI and VII and actual and calculated hide prices on Chart VIII.

Stocks of hides and leather are important, as shown by this analysis, but the most important factor is that of lower bracket income; and the next most important is the foreign exchange gradient. The study shows that when hides are valued on the basis of supply factors alone serious mistakes may be made, and that forecasting future income will return larger dividends than studying any other one factor.

VI. CONCLUSIONS

By use of new data on new orders, new indexes of visible stocks, the rate of change of bank deposits, the rate of change of foreign exchange, and a differential-integral mathematical expression representing speculative carry-over, it is possible to explain variations in general price indexes such as Fisher's index of industrial prices or Fisher's index of agricultural prices. The resulting calculated price indexes actually lead prices by several months. Here is provided a means for relating the pioneer classical studies of demand of Schultz,

[21] This method of analysis is due to the author's associate, Victor S. von Szeliski, who has co-operated in all the studies presented here.

CHART VI-A
Factors Affecting Hide Prices

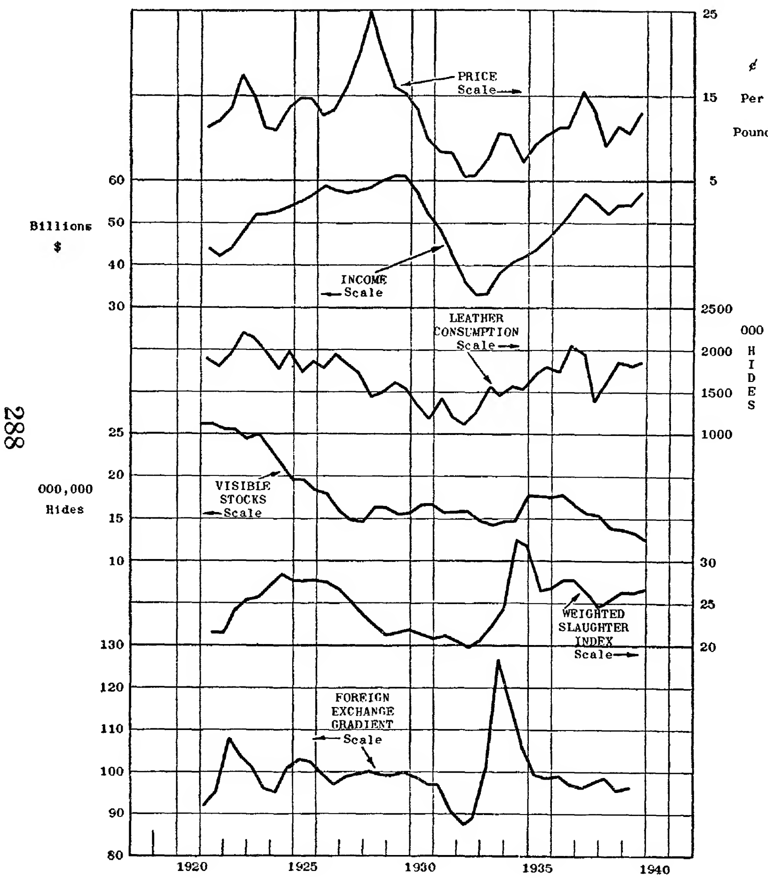


CHART VI-B
Factors Affecting Hide Prices

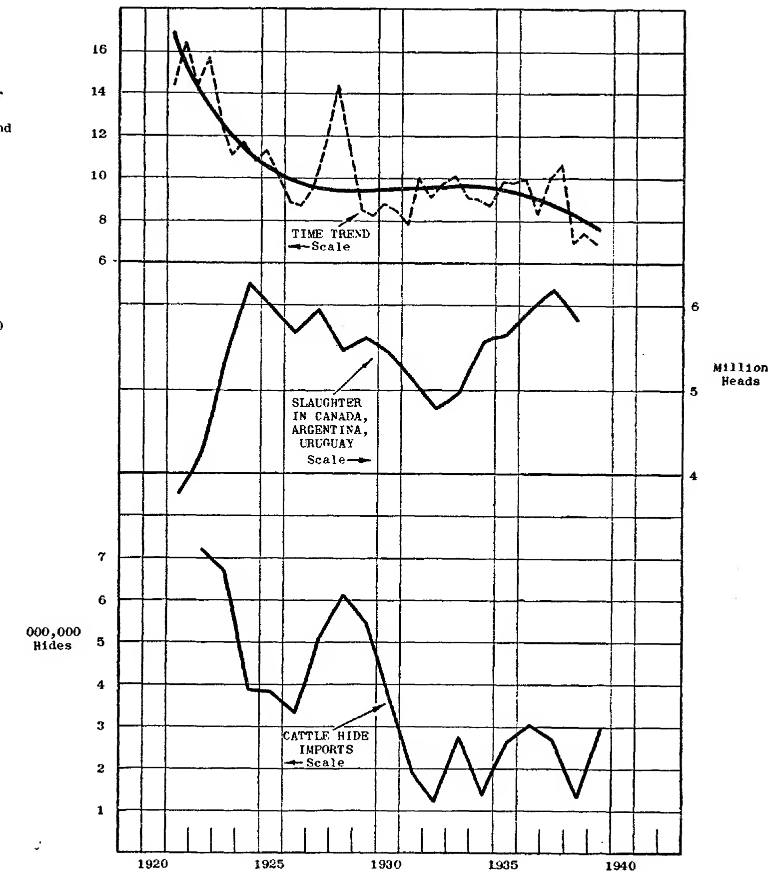


CHART VI-C
Hide Prices vs. Income

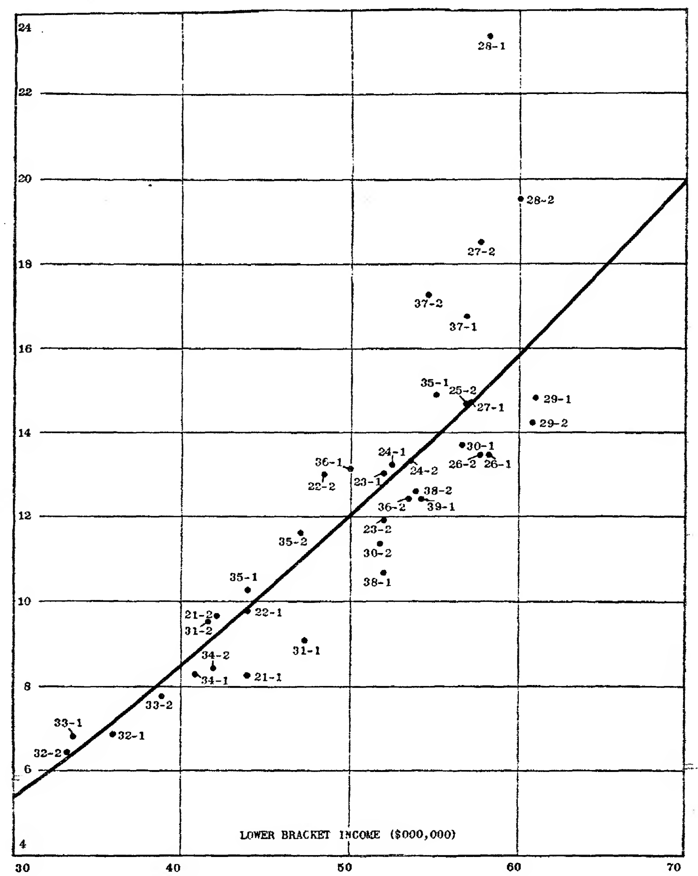


CHART VI-D
Calf and Kip Leather

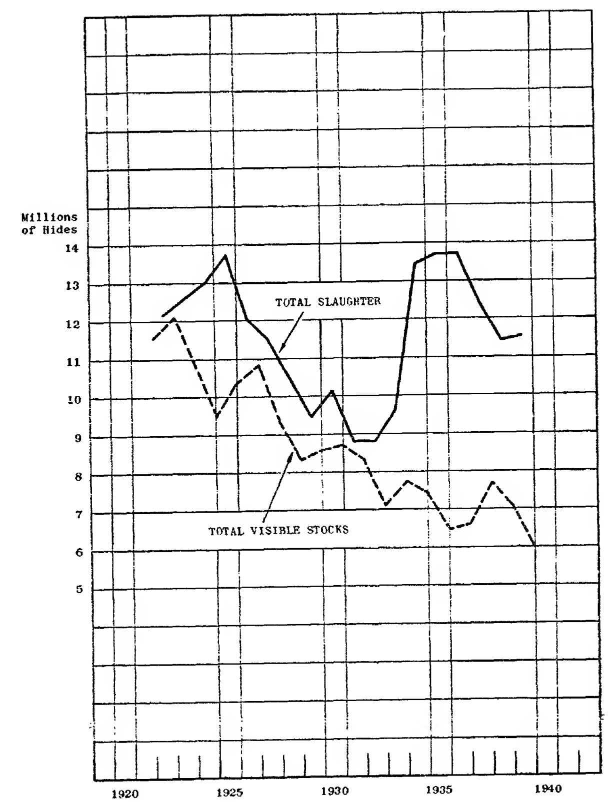


CHART VII-A

HIDE PRICES VS. TIME TREND

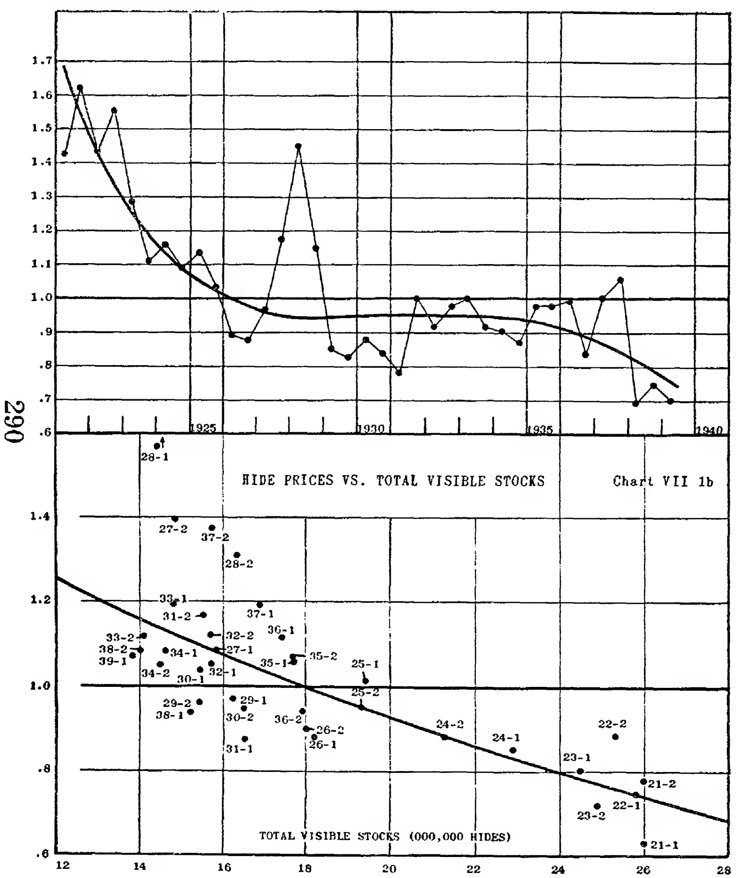


CHART VII-B

HIDE PRICES VS. LEATHER CONSUMPTION

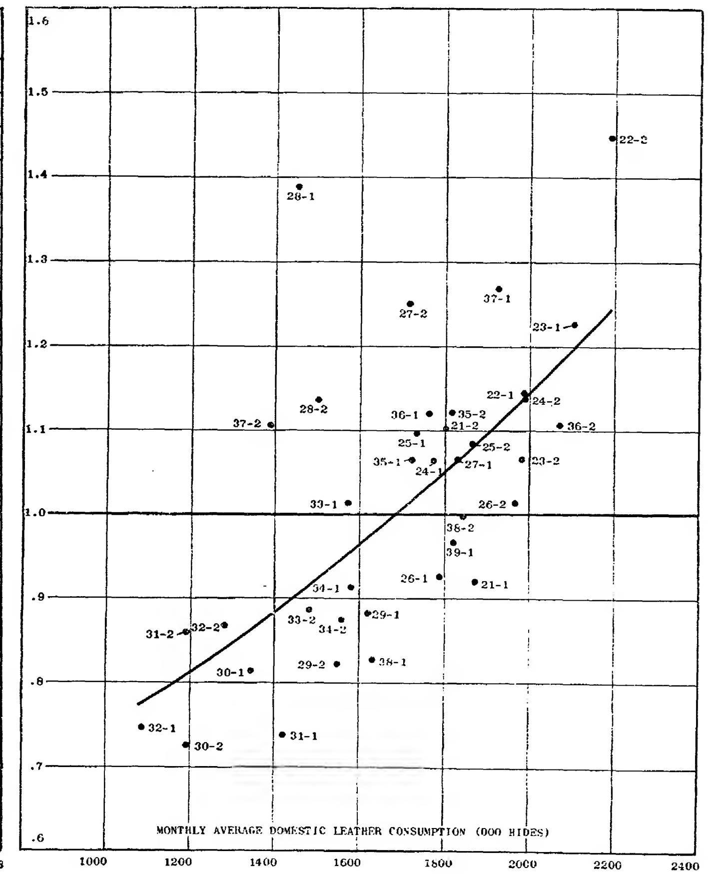


CHART VII-C
Hide Prices vs. Weighted Slaughter

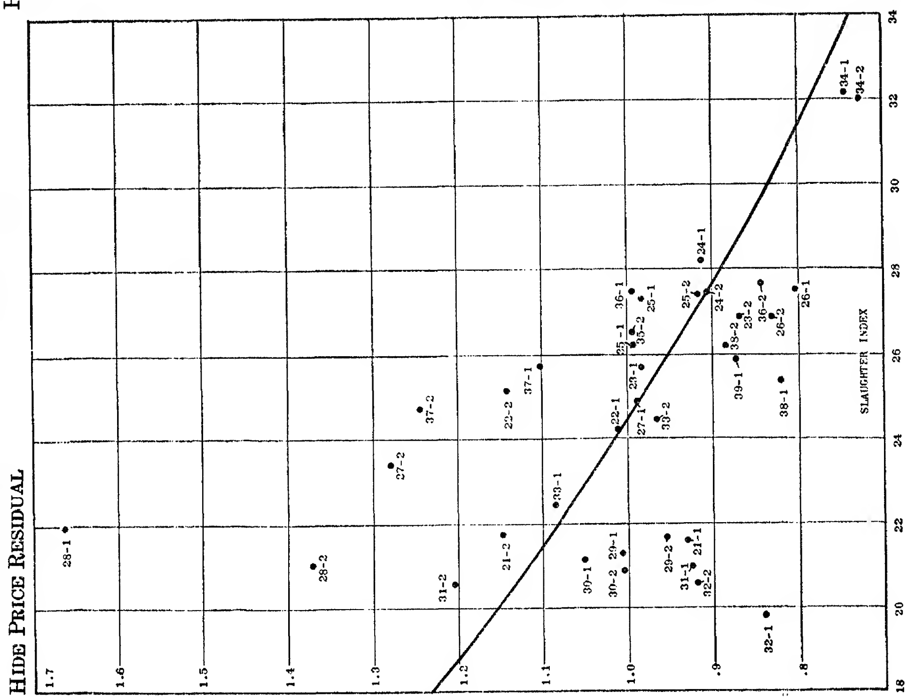


CHART VII-D
Hide Prices vs. Foreign Exchange Gradient

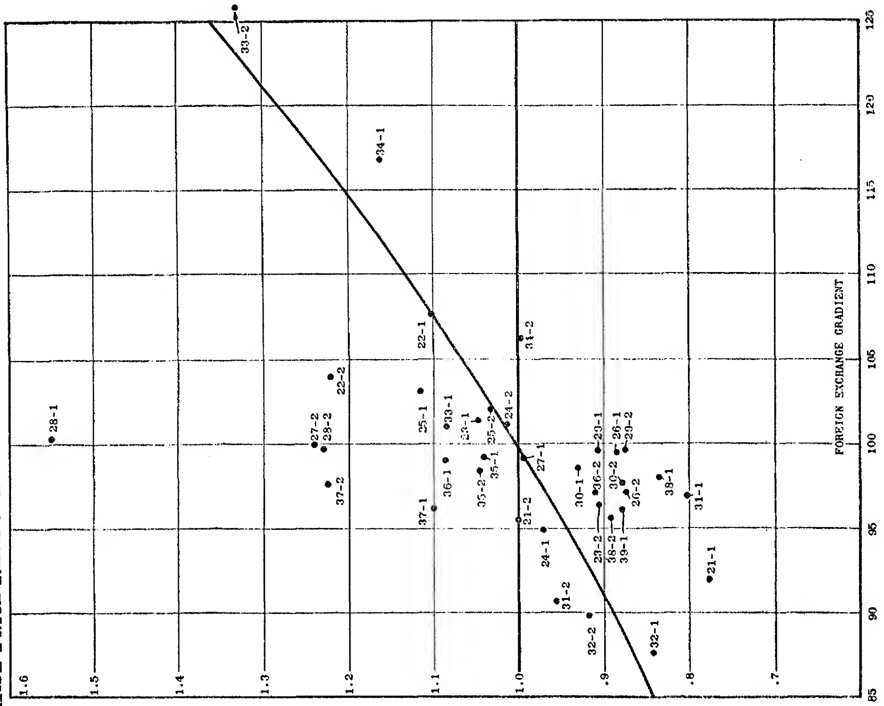
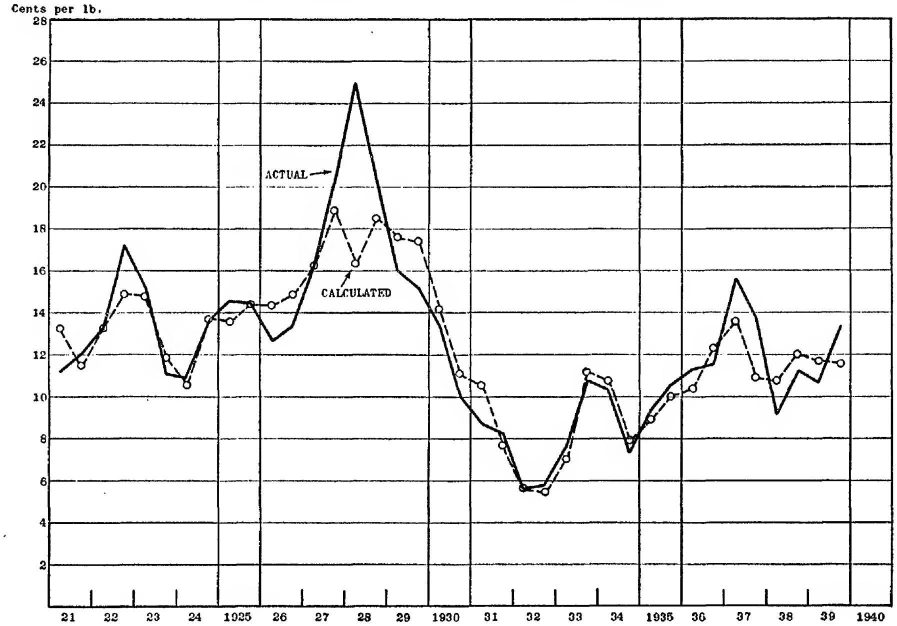


CHART VIII
HIDE PRICES—ACTUAL AND CALCULATED



Ezekiel, Working, and others, which amount to verification of "static" theory, to dynamic economic theory.

While most prices of individual commodities show price fluctuations corresponding to those of general indexes, superior understanding of the dynamics of individual prices will be attained by analyzing prices without using general "deflating" indexes. Such studies are herein presented for copper, wool, and hides.[22]

[22] Similar studies have been made by the Institute of Applied Econometrics for cottonseed oil, tallow, lard, cotton, lead, steel scrap, and lumber. Other studies are in progress.

PRINTED IN THE U.S.A.